DOCUMENT RESUME

ED 457 245 TM 033 372

AUTHOR	Vinovskis, Maris A.
TITLE	Overseeing the Nation's Report Card: The Creation and Evolution of the National Assessment Governing Board (NAGB).
INSTITUTION	National Assessment Governing Board, Washington, DC.
PUB DATE	2001-00-00
NOTE	92p.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Educational History; Elementary Secondary Education; National Competency Tests; Research Design; *Test Construction; Test Results; Test Use
IDENTIFIERS	*National Assessment Governing Board; National Assessment of Educational Progress; Standard Setting

ABSTRACT

This paper describes the creation of the National Assessment Governing Board (NAGB) in 1998 and examines the background characteristics of the Board members and their attendance at NAGB meetings. The staffing and financing of the NAGB and the relationship between the agency and the National Center for Education Statistics is also considered. The paper also explores two of the major issues addressed by the NAGB: the reporting of state-level National Assessment of Educational Progress (NAEP) data and the setting of student performance standards. The paper concludes with some observations about NAGB development and its functioning during the past 10 years and some recommendations for improvements. The NAGB and the NAEP have played an important role in telling how well U.S. children are doing in school and defining what expectations for them should be. The next step is to provide the effective education that students need to reach the goals set out for them. Perfecting the operation of the NAGB and the NAEP without addressing the need for better research and development in the areas of school improvement models and classroom practices makes little sense. (Contains 335 endnotes.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

TM

ED 457 245

Overseeing the Nation's REPORT CARD

The Creation and Evolution of the National Assessment Governing Board (NAGB)

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

R. Fields

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

TM033372

Maris A. Vinovskis

Department of History • Institute for Social Research
School of Public Policy • University of Michigan

naep

National Assessment Governing Board
U.S. Department of Education

Overseeing the Nation's REPORT CARD

The Creation and Evolution of the National Assessment Governing Board (NAGB)

Paper prepared for the National Assessment Governing Board (NAGB). The interpretations and views expressed are strictly those of the author and do not represent the views of NAGB or any other organization.

Maris A. Vinovskis

Department of History • Institute for Social Research • School of Public Policy
University of Michigan • November 19, 1998

THE NATION'S REPORT CARD naep

Acknowledgments

This analysis was commissioned by the National Assessment Governing Board (NAGB) and I am grateful for its support. NAGB's staff, especially Roy Truby, Lawrence Feinberg, and Sharif Shakrani, provided extraordinary help and always shared with me whatever published and unpublished documents I needed. Current and former members of the National Center for Educational Statistics (NCES), particularly Pascal Forgione, Jr., Emerson Elliott, and Gary Phillips, helped me better understand the National Assessment of Educational Progress (NAEP) and how it has changed over time. I owe a special debt to Chester Finn, who met with me on numerous occasions in Washington and thoughtfully read multiple drafts of my work.

Preliminary drafts of this analysis were shared with approximately 75 individuals—many of whom generously shared their personal knowledge of this subject, provided me with additional documents, or commented on my work. Among those individuals were Rolf Blank, Mary Blanton, Mary Lyn Bourque, Richard Boyd, Peggy Carr, Gregory Cizek, Joseph Conaty, Chris Cross, Mary Crovo, Ray Fields, Mary Frase, Arnold Goldstein, Steven Gorman, Richard Jaeger, H. Thomas James, Jack Jennings, Carl Kaestle, Michael Kane, Andrew Kolstad, Ellen Condliffe Lagemann, Jeffrey Mirel, Mark Musick, Michael Nettles, Larry Ogle, Martin Orland, Lois Peak, William Randall, Diane Ravitch, Jeffrey Rodamar, Edward Roeber, Bella Rosenberg, Michael Ross, Robert Rothman, Lee Schulman, Ramsay Selden, Daniel Stufflebeam, Stephen Swearingen, Suzanne Triplett, and Mary Ann Wilmer. Andris Vinovskis helped assemble the relevant research materials and Mary Vinovskis edited earlier drafts of the paper. At Aspen Systems Munira Mwalimu supervised the preparation of this report for publication and Ellen Burchill and Laura Mitchell King expertly copyedited the text. Graphic design and production expertise provided by Neeta Datt and Jennifer Baugh.

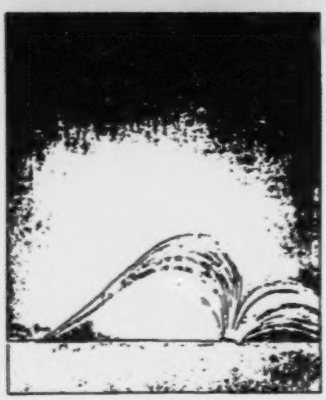


Table of Contents

Acknowledgments .iii

Introduction 1

I Early Efforts To Collect and Use Comparative Educational Data 3

II The Creation of the NAEP Tests in the 1960s and 1970s 5

III Improving NAEP in the 1980s and the Creation of NAGB 11

IV The NAGB Members 21

V Staffing and Financing NAGB 27

VI Interpreting NAGB's Authority and Responsibilities 29

VII Implementing State-Level NAEP 35

VIII Developing NAEP Performance Standards 41

IX Personal Observations about the Future Directions of NAEP and NAGB 59

Endnotes 69

Introduction

The tenth anniversary of the National Assessment Governing Board (NAGB) provides an opportunity to reflect on the agency's past as well as to reexamine some of its policies. NAGB was established to oversee the National Assessment of Educational Progress (NAEP) tests, which were created in the mid-1960s. Therefore, it will be useful to trace the development of the NAEP assessments to appreciate why it was thought necessary to establish NAGB in 1988.

After the analysis of NAGB's creation in 1988, this paper will investigate the background characteristics of the Board members and their attendance at NAGB meetings. The staffing and financing of NAGB and the relationship between the agency and the National Center for Education Statistics (NCES) will then be considered. The paper also will examine two of the major issues addressed by NAGB—the reporting of state-level NAEP data and the setting of student performance standards. The paper concludes with my personal observations about NAGB's development and functioning during the past ten years and some recommendations for future improvements.

Given the modest scope of this project, a number of other important issues must await future analysis. For example, the setting of test content frameworks and the debates over the types of background questions that should be gathered could have been investigated. This project was also unable to examine the advisability of adjusting NAEP scores to account for different student backgrounds and experiences or to consider the statistical techniques that should be employed in analyzing NAEP data. The total amount of monies spent on NAEP during the past three decades should be investigated, as well as its overall impact on educational reform in the United States. Given the limited time and resources available for this project and the disappointing lack of adequate secondary analyses, these topics could not be pursued in more detail. This study hopefully will provide a useful introduction to the history of NAGB and stimulate additional research in the near future.[1]

I Early Efforts To Collect and Use Comparative Educational Data

Colonial Americans, especially those in New England, were very interested in education, but initially chose to teach their own children and servants how to read. Yet parents increasingly wanted to send their children to local schools rather than teach them at home.[2] After the American Revolution, interest in education rose even more as political leaders and educators stressed the need for an educated citizenry in the new republic.[3] Large regional differences in education continued, with New England the leader in terms of white adult literacy and an extensive system of schooling. Even within educationally progressive states such as Massachusetts and Connecticut, there were sizable community disparities in the provision of formal educational opportunities (especially as rural areas failed to keep pace with the growth of schools in the larger cities).[4]

As schooling spread unevenly throughout the United States in the first half of the nineteenth century, reformers sought to create state education systems to persuade and coerce reluctant communities and parents to educate their children. Continued fear of any centralized government power, however, meant that few state school superintendents were given any real authority or power to control local education. Instead, most state education superintendents were limited to collecting statistical data from district schools and allowed to use that information only in their annual reports to encourage local school committees to improve their educational offerings.[5]

The extensive use of educational statistics and examples by nineteenth-century advocates to reform education probably seems somewhat simplistic and naive to us today. Yet it was based on the widespread contemporary belief in the intrinsic value of numerical data and the power of simple comparisons among schools to change local practices. Nineteenth-century reformers had an abiding faith that the compilation and display of numerical data not only would reveal the inherent regularities in behavior, but also would suggest possible options for making changes. They believed that if policymakers and the public were presented with the appropriate comparative data on social reforms such as education, they would soon want to improve their own policies accordingly.[6] Although nineteenth-century educators usually displayed little understanding of or appreciation for rigorous social science research, most school reformers accepted the importance and utility of collecting and disseminating educational information and sharing proven practices with each other. Several unsuccessful efforts were made in the 1830s by private educational groups to collect such information at the national level.[7]

Despite several attempts in the first half of the nineteenth century, little headway was made to increase the involvement of the federal government in promoting state and local education. The relative neglect of schooling during the Civil War and the need to improve educational opportunities in the vanquished South after the war, however, lent additional support to the idea of creating a federal education agency. The ascendancy of the Republicans in Congress and the White House who favored more government assistance for some domestic programs also helped to pave the way for more federal government involvement in domestic affairs. The supporters of a strong federal educational presence eventually had to settle for a more modest agency than they had envisioned.[8] Representative James Garfield (R–OH) finally introduced the bill to create a Department of Education in 1867:

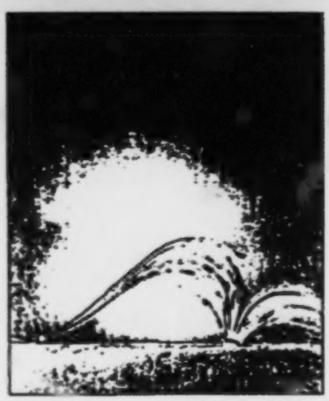
> *Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,* That there shall be established, at the city of Washington, a Department of Education, for the purpose of collecting such statistics and facts as shall show the condition and progress of education in the several States and Territories, and of diffusing such information respecting the organization and management of schools and school systems, and methods of teaching, as shall aid the people of the United States in the establishment and maintenance of efficient school systems, and otherwise promote the cause of education throughout the country.[9]

Although the promoters of the Department of Education had hoped for a much more active role for the agency than collecting and disseminating statistical information, the strong negative reactions against the poor administrative practices of the first Commissioner of Education, Henry Barnard, helped to doom those prospects in the short run. Barnard was forced to resign and the agency was demoted to a Bureau of Education within the Department of the Interior. John Eaton, Barnard's successor, focused the agency more narrowly on gathering and disseminating educational data; yet he managed to expand its staff over the next fifteen years from two to thirty-eight employees.[10]

By the early decades of the twentieth century, the Bureau's collection and analysis of educational data and information had improved considerably. States and localities were providing more uniform educational data and the results were published biennially. Additional information about innovative educational practices was gathered more systematically and analyzed by the staff. But these tasks now composed only a minuscule part of the overall budget as the agency acquired new responsibilities, such as administrating educational and relief programs in Alaska (which made up nearly sixty percent of the Bureau of Education's budget in 1920).[11]

Although educational statistics continued to be collected and used in the first half of the twentieth century, there was a growing recognition of their limitations in promoting educational reforms by themselves. As scientific research on children and schools progressed, educators and reformers placed more emphasis on supporting research studies than on just collecting and disseminating statistics. The Bureau of Education did try to use comparative statistics as a spur to educational improvement by classifying and rank-ordering colleges and universities, but the resultant political furor ended such efforts decisively.[12]

Repeated attempts to increase the role of the federal government in education in the 1930s and 1940s failed, but the now renamed U.S. Office of Education (USOE) expanded its activities during World War II. After the war, support for educational research and statistics continued to lag far behind those of the other behavioral and social sciences.[13] But the launching of Sputnik by the Soviets in October 1957 led to a substantial increase in the federal role in education. Although some federal programs for K–12 education like PSSC Physics were enhanced, most members of Congress still were not prepared for a larger federal role in elementary and secondary education. Instead, the legislators focused on providing more funding for higher education. The passage of the National Defense Education Act (NDEA) (P.L. 85–864) in September 1958 expanded federal support for graduate education and provided additional funds for the existing cooperative research program.[14]



II The Creation of the NAEP Tests in the 1960s and 1970s

Federal involvement in education grew rapidly during the 1960s. President John F. Kennedy sought to expand the federal role in the early 1960s, but failed to secure the necessary congressional support to enact his proposed education programs. Following his assassination in 1963, public sympathy for Kennedy, together with a weak opponent in Barry Goldwater, contributed to a landslide victory for his successor, Lyndon Johnson. The 1964 election also brought a more Democratic Congress to Washington, which Johnson was able to persuade to pass the historic Elementary and Secondary Education Act (ESEA) of 1965. This act provided federal aid for disadvantaged students and more monies for federal research and development.[15]

Part of the expansion of federal involvement in education was the planning and development of a national student assessment system during the 1960s. Concerns about ways to assess students reflected in part the growing interest in accountability in government during the Kennedy administration.[16] The federal official most responsible for the creation of this assessment was Francis Keppel, the U.S. Commissioner of Education from 1962 to 1965. Keppel, a former dean of the Harvard School of Education, lamented the lack of information about the academic achievement of American students:

> It became clear that American education had not yet faced up to the question of how to determine the quality of academic performance in the schools. There was a lack of information. Without a reporting system that alerted state or federal authorities to the need for support to shore up educational weakness, programs had to be devised on the basis of social and economic data.... Economic reports existed on family needs, but no data existed to supply similar facts on the quality and condition of what children learned. The nation could find out about school buildings or discover how many years children stay in school; it had no satisfactory way of assessing whether the time spent in school was effective.[17]

Keppel was careful to call for the assessment of students in terms of his responsibilities as the commissioner of education to collect and disseminate educational information to Congress. According to the justifications for the establishment of the agency in 1867, no one could deny that the commissioner of education had that authority and responsibility. But many educators doubted that this was Keppel's sole or even primary motivation for seeking to establish a national assessment of students. Instead, they feared that Keppel was simply using that rationale to create an assessment instrument to increase federal power over state and local education and perhaps even move toward a national curriculum. Although Keppel denied that he had any ulterior motives in establishing a national system for assessing students, there are indications that he was less interested in narrowly discharging his duties as the commissioner of education than in using the federal government to spur overall K–12 educational development.[18]

In mid-1963, recognizing the serious technical and political difficulties involved in creating a national student assessment system, Keppel called on Ralph W. Tyler, a psychologist and the nation's most prominent educational evaluator, for assistance.[19] With funding from the Carnegie Corporation, preliminary conferences were held in September and December 1963 and an Exploratory Committee on Assessing the Progress of Education (ECAPE) was created in June 1964 with Tyler as chair.[20] Based on his own assessment experiences and suggestions from other experts, Tyler proposed periodically assessing a small sample of different students rather than trying to test nationally all students. Faced with strong opposition from several major educational associations, Tyler tried to allay the fears of the critics of the proposed national assessment:

> This project is encountering some difficulties in getting itself understood. It is being confused with a nationwide, individual testing program, and several common fears are expressed by those who make this confusion. They note that tests used in a school influence the direction and amount of effort of pupils and teachers. In this way, if national tests do not reflect the local educational objectives, pupils and teachers are deflected from their work. This criticism does not apply to the assessment project because no individual student or teacher can make a showing. No student will take more than a small fraction of the exercises. No scores will be obtained on his performance. He will not be assessed at any later time and can gain no desired end, like admission to college or a scholarship.
>
> A second fear is that such an assessment enables the federal government to control the curriculum. This is also a misunderstanding. The objectives to be assessed are those which are accepted by teachers and curriculum specialists as goals toward which they work. They have been reviewed by lay leaders throughout the country so as to include only aims deemed important by public-spirited citizens. This project will report on the extent to which children, youth, and adults are learning things considered important by both professional school people and the informed public.
>
> A third fear is sometimes raised that this project would stultify the curriculum by not allowing changes over the years in instructional methods and educational goals. It should be made clear that the project will assess what children, youth, and adults have learned, not how they have learned it. Hence, the assessment is not dependent upon any particular instructional methods.[21]

As criticisms of the proposed student assessments mounted, Keppel, Tyler, and other supporters retreated from the idea that the results should ever be compiled to coerce states or local schools to improve their education. Keppel and participants in the early Carnegie-funded workshops had expected that the outcomes from the student assessments would be collected at the state and perhaps even the local levels—thereby encouraging state and local officials to reform their schools to remain competitive with other areas. Moreover, federal officials could have used the state-level data to decide how to allocate federal education dollars.[22]

Several influential educational associations were opposed to any student assessment data being collected and released at the state level because they feared that the results would be used to make improper and harmful comparisons. Organizations such as the American Association of School Administrators (AASA) initially were so opposed to the plans that they urged their members not even to participate in the pilot projects for the proposed assessments. And the president of the National Council of English Teachers admonished teachers "to fight tooth and nail to prevent a proposed plan to measure the quality of American education."[23] As a result, Tyler and the other members of ECAPE were forced to abandon their plans for reporting the data at the state level. Appearing on a panel at the AASA Annual Meeting in February 1966, Tyler assured the superintendents that the smallest geographic unit for which the results would be reported was one of four regions:

> This emphasis, for example, on no smaller geographical region than the regions represented by the four in the United States—Northwest, Southeast, West and Far West—is one means of ensuring that we are not talking about comparing one state with another. We

BEST COPY AVAILABLE

> are not talking about comparing one kind of community with another. My own belief is that whatever may be the need and the desires of the Congress in trying to assess their responsibilities, our concern is with the assessment of our educational development as a whole, which includes children who may have been educated in parochial schools, private schools, as well as public schools, who may have been at home or out of school altogether.[24]

Promising not to release the results at the state level helped to calm the fears of some critics, but others still remained suspicious of the real motives of the proponents of the student assessments. Their fears were somewhat alleviated when George Brain, an AASA official, was elected chair of ECAPE. The test supporters in 1969 prudently transferred the administration of the student assessments to the Education Commission of the States (ECS)—a recently formed compact of states that could be trusted not to infringe on the rights of its members.[25] As a result, much of the hostility toward the national assessment of students gradually disappeared and the focus turned to developing and implementing the proposed assessments.[26]

Intermittent work on the proposed student assessments had been proceeding since 1963, with substantial private funding provided mainly by the Carnegie Corporation.[27] Several corporations, expert in evaluation and test development, helped to develop appropriate prototypes for those assessments.[28] The entire assessment development process took much longer than had been planned, largely due to the unanticipated difficulties in constructing such relatively new and novel instruments in ten subject areas, for four age groups (including young adults), and reflecting different levels of student competence.[29]

The development and refinement of matrix sampling in the 1960s and 1970s made the national assessment technically feasible because it provided a statistical means of asking each student only a few items, but still obtaining sufficient information on a much larger number of questions for subgroups of the population. In addition, the procedure allowed for compilation of accurate aggregate data, but did not provide reliable or usable individual-level results—thereby relieving some of the concerns of educators and parents who feared that particular children might be judged or compared against others on the basis of a national assessment.[30] Although the sampling frame was designed to provide information at four regional and seven general types of community levels, it avoided providing any state-level results or particular community findings.[31] The decision to avoid state-level, specific community-level, or individual-level data had been necessitated by the staunch opposition of several educational associations and some parents and teachers, but it also deprived the assessments of much of their practical usefulness for educational decisionmakers at the state and local levels.

Information on the personal characteristics of the students included data on their age, sex, race, and the educational level of their parents.[32] No attempt was made to gather data on the income level of the parents. This significantly limited the ability of researchers to provide policymakers with analyses that could take into consideration family income, even though the level of poverty was a major issue in most federal educational programs (such as ESEA).[33]

Given the analytic compromises necessary to establish the assessment program, especially the eventual decisions not to gather individual-level data or to aggregate the results at the local school or state levels, some critics wondered whether the resultant package was very helpful for educators and policymakers. For example, Martin Katzman and Ronald Rosen, who favored the idea of a large-scale student assessment in principle, questioned in 1970 whether the actual program would be useful in practice:

> One gets the overall impression that CAPE [Committee on Assessing the Progress of Education], in its attention to details of statistical validity, simplicity of administration, and use of a quasi-scientific approach, has lost sight of its major aims. It may seem amazing that such a large undertaking could go so far astray, but this becomes understandable when viewed in the perspective of its growth. Overreacting to early opposition, CAPE has evolved to a point of considerable ambivalence with respect to its original purpose of improving educational decision-making at the local, state, and

> federal level. It is quite clear that National Assessment will provide little information on the policy issues of the day—the effects of segregation, the effects of decentralization, the effects of resource or curriculum shifts. Nevertheless, considerable lip service is paid to the notion that assessment will improve policy....
>
> The National Educational Assessment Program as it stands today can be criticized on several grounds: 1) measuring questionable educational outcomes with questionable techniques; 2) classifying student subpopulations on largely irrelevant dimensions and/or insufficient detail; 3) neglecting to collect any information on school characteristics which would identify policy-performance relationships. In principle all of these shortcomings can be remedied; however, the institutions for administrating the program make such remedy unlikely. We question whether the budget for the program might be shifted to better forms of educational research.[34]

Others, like Galen Saylor, who had not favored a national student assessment in the first place, continued to object to the project and suggested that the money should be distributed to the states so that they could do their own evaluations, which he believed would be more helpful and useful to educators and policymakers:

> I would strongly urge that, instead of this farflung national assessment project, we begin developing in our state and local school systems some comprehensive programs of evaluation. It is from such evaluations that we can gather evidence of help to local boards of education, administrative staffs, and teachers interested in undertaking reforms, modifying existing programs, and developing the kinds of new programs that would assure the children and youth of the community an improved education. I would advocate, therefore, that Congress make large sums of money available to the states for assisting local districts in undertaking expert evaluations conducted by specialists in the field. Our present methods of evaluation are often inadequate, invalid, or inconsequential; but this is not to say that we need a program of national assessment. If we use available resources to improve evaluative programs at the local level, we can use the information derived in revising school programs and improving instruction.[35]

ECS assumed responsibility for directing and management of assessments in June 1969. USOE was expected to provide the long-term funding because it was anticipated that the private foundations soon would terminate their financial support. CAPE became an advisory committee and ECS set up a Project Policy Board to oversee the undertaking. The entire project was renamed the National Assessment of Educational Progress (NAEP).[36]

Prior to 1968 most of the funding for the assessment project came from the Carnegie Foundation. The federal government provided $372,358 in 1968, to which Carnegie added $1 million; in 1969 the federal contribution rose to $1 million and the Carnegie and Ford foundations gave $910,000. The following year USOE furnished $2.4 million and Carnegie made its last contribution of $350,000. In 1972 the federal government provided the entire $4.5 million. Within a short time, the funding for NAEP had shifted entirely from private sources to the federal government.[37]

Initially there had been considerable concern that if the federal government provided the funding for the assessment, NAEP would lose its independence and autonomy. But given the high and continuing costs associated with NAEP, there was little choice but to rely on the financial support of USOE. Moreover, ECS, which provided the policy oversight of NAEP, had been assured in 1969 that USOE would provide funding but not interfere with the policy or analytic aspects of the assessment. USOE initially maintained its part of the bargain, but then reneged in the early 1970s as Congress put more pressure on the agency to reduce its budget and monitor its grants and contracts more closely. Sydney P. Marland, Commissioner of Education, first transferred the monitoring of NAEP from the National Center for Educational Research and Development (NCERD) to the National Center for Educational Statistics (NCES)—thereby moving oversight of the program from a general educational research program to one solely concerned with collecting and analyzing data. He also converted NAEP funding from a grant to a contract and subjected the assessment to the normal scrutiny of the agency. The autonomy and independence that had been so carefully developed and protected for NAEP disapeared, despite earlier assurances to the contrary. As John Evans wrote on behalf of USOE to the ECS Policy Committee in October 1973:

> [W]e are increasingly conscious of the accountability we bear for this project and the substantial funds supporting it, and we have concluded that the nearly total independence which has characterized the NAEP grant thus far is not a satisfactory type of relationship for us to insure that the work of the NAEP is maximally policy relevant. Accordingly, we have decided, as I thought Sid Marland made unmistakably clear...to change the relationship towards one in which there would be more direction from the Office of Education....
>
> This...will involve changing the procurement instrument from a grant to a contract, specifying in the contract the major tasks and activities to be carried out, and requiring approval by OE of the major directions, activities, and products.[38]

As USOE played a larger role in the oversight of NAEP, additional questions were raised about the policy relevance of the assessments. Ironically, the decision to avoid compiling data at the state or local levels, which had been seen as essential for securing the cooperation of teachers and educational associations, now made the results from NAEP less useful to state and federal officials. Although there was some disagreement, especially from the original proponents of national student assessment, many observers in the 1970s continued to complain about the lack of policy relevance for much of the NAEP results. For example, in an overview of educational research and development in the 1960s through the mid-1970s, Richard Dershimer concluded: "Of what value was this national assessment through these years to the policy shapers in the federal government? Not much."[39] Similarly, a U.S. General Accounting Office (GAO) analysis of NAEP in 1976 concluded that its results should be made more useful to policymakers.[40] And an analysis of state legislators in the 1970s revealed that many were unaware of NAEP or did not use it much in their deliberations.[41] While the staff of NAEP had tried hard to be more policy relevant in the 1970s (and some later observers argued that they had succeeded more than had been realized at the time), the general impression among policymakers and educators in the 1970s was that NAEP was not particularly helpful to those in decisionmaking situations.[42]

The organization and oversight of NAEP was altered in 1978 when Congress enacted Public Law 95–561, which transferred the program to the National Institute of Education (NIE) and called for it to be either a grant or a cooperative agreement with a nonprofit education association. The legislation also created a 17-member Assessment Policy Committee that included two representatives of business and industry, three from the general public, four classroom teachers, two state legislators, two school district superintendents, one state governor, one chair of a state board of education, one chair of a local school board, and one chief state school officer. The Assessment Policy Committee was to be chosen by the contractor and was to be responsible for the design of NAEP as well as of the studies to evaluate its validity, effectiveness, and utilization.[43] Congress clarified the responsibility of the Assessment Policy Committee in 1984 to include information about the background materials as well.[44]

III Improving NAEP in the 1980s and the Creation of NAGB

The 1980s saw major changes in education and politics in the United States and a reorganization of the governance of the National Assessment of Educational Progress (NAEP). The landslide election of Ronald Reagan as president and the narrow, unexpected victory of Republicans in the U.S. Senate in November 1980 led to challenges to many of the existing federal domestic programs. Reagan and many of his more conservative Republican allies hoped to eliminate the recently established U.S. Department of Education. They failed because of unexpectedly strong, bipartisan congressional opposition to abolishing the department; the decision of the Reagan administration to concentrate initially on its other priorities; and the lack of enthusiasm and support for their project from the new Secretary of Education Terrell Bell.[45]

Although some conservative Republicans tried to dismantle the Department of Education and eliminate most federal involvement in education, in principle they were not opposed to the federal data collection and dissemination functions that had been established more than a century earlier. As long as the federal government did not try to use NAEP to regulate or coerce states and local school districts, most Republicans were ready to continue their support of that project.[46] Funding for NAEP, however, had already diminished considerably from a high of $6 million in fiscal year (FY) 1973 to $4.3 million in FY 1979. The following year appropriations dropped to $3.9 million, where they remained for four years.[47] Compared with other, more drastic cuts in the Department of Education in the early 1980s (especially in areas such as research), NAEP fared remarkably well, with less than a 10-percent reduction from the late 1970s.[48]

Interest in educational reform waned somewhat during the 1970s, but gained fresh momentum in the early 1980s as a series of reports detailed the "abysmal" state of schooling in America.[49] The most famous and influential document was the widely circulated report, *A Nation at Risk*, which challenged Americans in 1983 to return to the basics in education and to focus attention on student academic achievements.[50] The public reaction to the report was so strong and positive that Reagan decided to participate personally in the regional discussions of *A Nation at Risk* as a major part of his reelection campaign.[51]

Building on the public success of *A Nation at Risk*, in 1984 Secretary Bell embraced the idea of using a large wall chart to display the comparative educational progress of each state. In looking back on the development and use of the wall chart, the staff who developed and implemented that project commented on its impact:

> The wall chart has become the focus of considerable attention and controversy. Some analysts see state-by-state comparisons as filling a void in our statistical knowledge, enabling states and their residents to gauge for the first time the quality of their education. Others see this information as statistically flawed and providing little guidance to improve the system; worse yet, they say, the measures may mislead, sending reform efforts off in the wrong direction.
>
> We believe that the publication of the wall chart, with its acknowledged flaws, has helped validate state-by-state comparisons as a means of holding state and local school systems accountable for education. In fact, of all of the lessons learned from the wall chart, the most important has been establishing this validity.[52]

Annual updates of the wall chart provided Secretary Bell and his successors with an opportunity to applaud or criticize the educational achievements of the states.[53]

One of the major shortcomings of the wall chart was the lack of suitable state-level student achievement information. The Department of Education had used ACT or SAT scores, but these indices were roundly criticized by educators, who challenged their representativeness due to the absence of data on noncollege-bound youth and the noncomparability of those indices across different states (largely due to varying student participation rates among the states).[54] Yet the debates over the quality of the data for the wall chart and its apparent value for federal and state policymakers provided an enticing preview of how state-level NAEP results might be used if they became available.[55] Gerald N. Tirozzi, Commissioner of Education in Connecticut, observed that "the wall chart was just the beginning of what's to come. And I would rather have accurate, appropriate, and fair measures of comparison than biased, distorted, and inaccurate ones."[56]

Most state superintendents of education had been hostile to the idea of compiling and releasing state-level student assessments in the late 1960s and early 1970s. But gradually state officials became more interested in gathering their own student assessment data and some even wanted to make their tests more comparable to the national NAEP examinations. A few states developed their own student assessments in the 1970s—often with some technical assistance from the NAEP staff. This interest in state-level assessments continued to grow in the early 1980s.[57] In 1984 the NAEP's Assessment Policy Committee voted 19 to 2 to help states and local areas compare their own student assessments to the national NAEP.[58] Although some disagreement over the advisability of reporting state-level NAEP results still persisted, several states instituted their own student assessments.[59]

One of the major leaders in the movement for state-level assessments was the Southern Regional Educational Board (SREB). At its annual meeting in 1984, several state governors called for improvements in measuring educational progress. Governor Lamar Alexander of Tennessee remarked that "it's virtually impossible for me to persuade the taxpayers to give up another penny unless I can show them results."[60] Governor Bill Clinton of Arkansas agreed and said that comparing student achievement to a national norm would stimulate "competition in the best sense" and encourage school improvement.[61] Eight southern states in 1986 began a three-year test of a sample of their students using NAEP reading and/or writing achievement tests.[62]

In 1984, by a narrow vote of 20 to 19, the Council of Chief State School Officers (CCSSO) also approved plans for cross-state comparisons.[63] As a result, rather than being perceived as a threat to the well-being of the states, some governors and state legislatures were welcoming calls for the compilation and dissemination of state-level NAEP results.

At the same time that interest in state-level NAEP information increased, other major changes were occurring that encouraged a revision of the federal data-gathering system. The existing design and administration of NAEP was strongly criticized by former Labor Secretary Willard Wirtz and by Archie E. Lapointe, the future ETS director of NAEP, in a major study released in early 1982. Although the authors praised NAEP in principle, they felt it was "underdeveloped" and "underused"; Wirtz and Lapointe also believed that NAEP had "apparently negligible influence" on policymakers or teachers. They lamented the lack of adequate funding for NAEP and suggested that the program be eliminated altogether unless adequate financial support could be found.[64]

The next five-year NAEP contract was scheduled for renewal in 1983. The Educational Testing Service (ETS) won that competition and replaced the Education Commission of the States (ECS), which had managed the program since 1969, in large part because ETS promised to introduce more sophisticated statistical procedures and to make the results more useful to policymakers.[65] NAEP was redesigned in the early 1980s to cover four subject areas—reading, writing, math, and science—on a more frequent and regular schedule. In addition to the traditional assessments of 9-, 13-, and 17-year-olds, children in grades 3, 7, and 11 were to be examined. Improved matrix sampling of test items allowed for more rigorous analyses of the relationship between students' background information and their assessment scores. Finally, introduction of nonlinear scaling methods for data reporting allowed the clustering of related items.[66]

As the management of NAEP shifted from the state-oriented ECS to the nonprofit ETS, concerns about the governance structure of the enterprise surfaced—especially as it became clear that the Reagan administration planned to continue federal support for NAEP but wanted to make sure it would reflect state and local education interests. Denis P. Doyle, director of Education Policy Studies at the American Enterprise Institute, recommended in 1983:[67]

> A 15-member governing board, representing the natural clients of the NAEP, should be established. These natural clients are state governments, local education authorities, and the federal government—in that order.... The 12 state and local members should serve staggered four-year terms and should be removed only for cause. The voting members representing the federal government should serve by virtue of their federal position.[68]

Doyle was anxious to prevent any special interest groups from directly controlling NAEP, and did not want any slots for their members on the governing board (the ETS-appointed Assessment Policy Committee had educators, policymakers, and lay people on it). Although he expected the governing board to identify the important policy issues to be addressed, he also saw the need for a technical advisory board.[69]

These discussions about the future and nature of NAEP were occurring at the same time that scholars and the National Center for Education Statistics (NCES) were reviewing federal educational data-gathering operations in general. In 1984 the National Academy of Sciences (NAS) was commissioned by the Department of Education's Office of Educational Research and Improvement (OERI) to undertake a thorough review of NCES. The NAS panel issued an unusually harsh condemnation of the poor quality of data collection and dissemination by NCES:

> We wish to emphasize the seriousness with which we view the center's problems. We believe that there can be no defense for allowing the center to continue as it has for all too long.... Without strong and continuous commitment and demonstrated determination to undertake wide-ranging actions to change both the image and the reality of the center, we are unanimous in our conviction that serious consideration should be given to the more drastic alternative of abolishing the center and finding other means to obtain and disseminate education data....
>
> We emphasize strongly, however, that we believe the preferred course of action is to begin the process of improvement. As we have noted, the center's problems are long-standing and pervasive, but if faced openly they can in time, be overcome.[70]

Assistant Secretary Chester Finn and Emerson Elliott, the future first Commissioner of Education Statistics, stepped forward and provided the support needed to rescue the agency.[71] Yet in 1986, when the NAS report had just been issued and a new panel to investigate NAEP was being established, it was not clear whether NCES would be salvaged.

OVERSEEING THE NATION'S REPORT CARD

It was in this climate of heightened public concern about education, a growing perception of the need for better state-level student data, and efforts to reorganize NCES that Secretary of Education William J. Bennett in May 1986 formed a distinguished 22-member NAEP study group that was headed by Tennessee Governor Lamar Alexander (who was also chair of the National Governors' Association) and H. Thomas James (former president of the Spencer Foundation). The study group included individuals such as Hillary Rodham Clinton, Linda Darling-Hammond, Pascal Forgione, Bill Honig, Francis Keppel, and Michael Kirst. The Alexander-James study group transmitted its report to the Department of Education in January 1987. The study group acknowledged the value of NAEP, but then criticized the lack of state-level NAEP data:[72]

> But NAEP has a serious weakness, and this must be identified here at the outset, for correcting it is our Study Group's most important recommendation. The weakness is that while providing excellent information on what our children know and can do, it provides it only for the nation *as a whole*, and for a few large regions of the country. Whole-nation information is of course useful when we wish to gauge the performance of our children against that of children in other countries, whether rivals or allies. But in the United States education is a *state* responsibility, and it is against the performance of children closer to home that we want and need to compare the performance of our youngsters....
>
> If we think of NAEP as a weather map, today's assessment is designed to provide temperature, barometric pressure, and precipitation levels only for the United States as a whole and for a few large regions within it (the Midwest, for example), regions that are essentially meaningless for education matters. We propose, instead, a much expanded weather map that will not only provide such information for the whole country, but will also provide it for every state within it—and do so in such a way that a state or locality can readily produce similar data at the community or even neighborhood level. These data in turn can be compared with data from other communities, the entire state, or the nation, both now and over time.[73]

The Alexander-James study group questioned the narrow range of subjects that NAEP was covering—due mainly to the lack of adequate funding. Instead, they said:

> We urge regular assessment of reading, writing, and literacy; mathematics, science, and technology; and history, geography, and civics. Other skills and subject domains should from time to time be included. In every instance, the assessment instruments should examine acquisition of pertinent "higher-order" skills as well as basic skills, knowledge, and concepts.[74]

Although the Alexander-James study group endorsed compiling data by student age, it also wanted more attention given to collecting information for the key transition grades:

> As in the past, the nation's report card should continue to gather information on children aged nine, thirteen, and seventeen, but grade-level samples should be changed from the present grades 3, 7, and 11 to the more important "transition" grades of 4, 8, and 12. In addition, out-of-school seventeen-year-olds should be included and, in the assessment of literacy, older age groups should be included as well. By making these changes, we will regularly gather vital data about two of the most important issues in American education today: dropouts and adult literacy.[75]

The study group discussed NAEP's recent extension of the gathering of background and school variables to include such items as measures of students' homework and television watching as well as school-level information about principals and teachers. In its discussions of the estimated costs of NAEP, the Alexander-James study group indicated its interest in expanding the nature and quality of parental background information by suggesting a separate questionnaire for parents of fourth-grade children (as these children probably were too young to provide reliable data on their families).[76] At the same time, the study group cautioned against gathering excessive school-level information unless there were reasons to believe that it may have a significant impact on student achievement.[77]

The Alexander-James study group called for the creation of a new Educational Assessment Council (EAC) to oversee the redesign of NAEP and proposed that EAC be provided permanent staff.[78] EAC members would serve five-year terms and include current and former educators, state or local school officials, testing and measurement experts, researchers, and

BEST COPY AVAILABLE

curriculum specialists. The secretary of education would appoint the members. A permanent standing committee would be established to nominate potential future members.[79]

The study group expected the federal government to fund and oversee most of the work of the contractor selected to conduct the assessments.[80] Given the increased magnitude of the new assessments envisioned by the study group, the expected annual cost of NAEP would rise from about $4 million to $26 million. Much of that increase (approximately $13.5 million) would be used to compile and analyze data at the state level. The estimated cost of the EAC and its professional staff was $2.5 million.[81]

The Department of Education asked the National Academy of Education (NAE) to review the Alexander-James study group's report and to publish and distribute the report and their comments on it. NAE appointed a six-member committee under the leadership of Robert Glaser of the University of Pittsburgh. The NAE committee praised several of the key recommendations of the Alexander-James study group,[82] but questioned whether NAEP alone could provide the information and studies needed for school reforms in the United States:

> What is less clear in the panel report is how NAEP data will actually link to school improvement efforts. Although NAEP can tell us a great deal about "how our schools are doing," it provides only limited and mostly indirect evidence about the factors contributing to these successes and failures. It is natural to suggest that NAEP data collection be expanded so as to shed more light on these casual linkages. Unfortunately, few such questions are well suited for examination within the current NAEP design.... In fact, this basic research is probably better pursued as a separate enterprise within the larger educational research community than as a small add-on to a large federal effort whose principal purpose is quite different.[83]

NAE expanded on its view of the limitations of the existing NAEP approach by recommending support for smaller, more intensive studies that would provide information about the schooling process.[84]

While the NAE committee did not oppose state-level NAEP assessments, it worried that the report by the Alexander-James study group overemphasized the importance and utility of this approach:

> We are concerned about the emphasis in the Alexander-James report on state-by-state comparisons of average test scores. Many factors influence the relative rankings of states, districts, and schools. Simple comparisons are ripe for abuse and are unlikely to inform meaningful school improvement efforts.
>
> State average scores on tests like the SAT have been much misused. Although the sampling technique proposed for NAEP will obviate many of these abuses, the ability of a state or locality to examine its progress over time is much more informative than the comparison with other states or localities at any one point in time. Because of the many variables contributing to the diversity of our educational institutions, among states and among localities, the simple ranking of geographic units by achievement levels is rarely informative. Not surprisingly, schools with greater resources and fewer problem students routinely fill the upper ranks. So what have we learned?[85]

Statistical adjustments could make the data more comparable, but they still provided little information about how to improve the schools.[86] NAE members feared that the high costs of state-by-state comparisons might preclude other, more worthy projects that could facilitate school improvements.[87]

In a suggestion that would lead to considerable debate, NAE recommended the development of several student performance levels rather than reporting results using arbitrary and hard-to-understand numerical score categories:

> We recommend that, to the maximal extent technically feasible, NAEP use descriptive classifications as its principal reporting scheme in future assessments. For each content area NAEP should articulate clear descriptions of performance levels, descriptions that might be analogous to such craft rankings as novice, journeyman, highly competent, and expert. Descriptions of this kind would be extremely useful to educators, parents, legislators, and an informed public.[88]

NAE applauded the idea of setting up a separate, relatively independent government board for NAEP. But it noted the ambiguity in the Alexander-James report about the relationship between the proposed EAC and the Department of Education—it was not clear whether the EAC's recommendations were binding or simply advisory. NAE hoped that the new governing board would be as independent as possible and urged that this matter be clarified immediately to prevent any future misunderstandings.[89]

The general response to the Alexander-James report was favorable, although still cautious about the long-term implications of the proposed changes. At a press conference announcing the release of the report in March 1987, Secretary Bennett embraced the recommendations and said, "I certainly intend to move forward with the legislation and to seek authorization to put an improved report card into the nation's hands."[90] CCSSO responded by proposing that NAEP develop state-level assessments in the *core* subjects of reading, writing, and literacy; mathematics, science, and technology; and history, geography, and civics. The council also recommended the establishment of an independent agency to oversee future assessments.[91] NAEP's Assessment Policy Committee also endorsed the plans for the expansion as proposed by the Alexander-James report, but expressed concern that since the new oversight group would be appointed by the secretary of education, it might lead to more federal control: "This change in governance, when combined with concerns expressed about the possible standardization of a system of state comparisons, may create an unintended impression of considerably increased federal influence over education."[92]

The legislative reorganization of NAEP became part of the larger reauthorization of the Elementary and Secondary Education Act (ESEA) of 1965 (P.L. 89-10). ESEA was last reauthorized in 1981 when the Reagan administration shifted more responsibility for remedial education programs to the states. The reauthorization of ESEA was debated in 1986 and 1987, and each chamber overwhelmingly passed its own version of the legislation by December 1987 (though the final reauthorization was not enacted until April 1988).[93] Because ESEA was the major federal compensatory education program, legislators in both the House and the Senate focused mainly on issues such as the targeting of federal funds to low-income areas or the need to support bilingual education. Relatively little attention was paid to the reauthorization of NAEP—in part because much of the discussion and debate about ESEA in the House had already concluded before the Alexander-James report was issued.[94]

The House bill (H.R. 5, *The School Improvement Act of 1987*), followed the earlier recommendations of NAS and focused on reorganizing NCES as a more independent statistical agency—something the Reagan administration opposed.[95] The House did call for the creation of a National Cooperative Education Statistics System within NCES that would produce and maintain comparable data (with states participating in this system on a voluntary basis).[96]

The House bill was relatively silent about the existing NAEP except for recommending that NAEP also compile longitudinal data on the achievement of students participating in the Chapter 1 program of ESEA.[97] At the same time, the House made it clear that it did not want the reorganized NCES to conduct evaluations of specific federal education programs:

> It is essential that the statistics identified to be collected and published by the National Center for Education Statistics stem from generic issues fundamental to understanding the nature of the education industry and its impact on the economy and society at the local, state and federal levels. Although the Committee expects that the Department of Education might seek advice on its responsibilities to evaluate and monitor federal education programs, the purpose of the National Center for Education Statistics is not to conduct evaluation of specific federal education programs. Fundamental to the trust the public has in the truthfulness of an agency's statistics is the belief that the data are not biased toward any particular ideology.[98]

The Senate focused more on improving NAEP, but paid little attention to NAS suggestions for the reorganization and increased independence of NCES. Under the leadership of Edward Kennedy (D-MA), the Senate incorporated most of the Alexander-James study group's recommendations in its ESEA reauthorization (S. 373). The Senate bill expanded the number of educational subjects to be assessed (reading, writing, mathematics, science, history, geography, and civics); called for gathering and reporting state-level data on a voluntary basis; created a 20-member National Assessment Governing Board (NAGB) to oversee NAEP; and authorized at least $11.5 million for FY 1989, $17.7 million for FY 1990, $17.9 million for FY 1991, and $19.6 million for FY 1992 and FY 1993.[99]

The Senate legislation directed the secretary of education to appoint NAGB members to staggered four-year terms; for each future vacant position the board would submit three nominations to the secretary.[100] NAGB membership was to be "balanced fairly in terms of geographical distribution and the points of view represented and that it exercises its independent judgment, free from inappropriate influences and special interests." The legislation specified that the twenty members would include individuals from specifically designated categories.[101]

According to the Senate bill, NAGB was to "design and supervise the conduct of the National Assessment." The board was to:

> select subject areas to be assessed; identify feasible achievement goals for each age and grade in each subject area to be tested under the National Assessment; develop assessment objectives; develop test specifications; design the methodology of the assessment; develop guidelines and standards for analysis plans and for reporting and disseminating results; develop standards and procedures for interstate, regional and national comparisons; and take appropriate actions needed to improve and the form and use of the National Assessment. The Board shall have final authority on the appropriateness of cognitive items.[102]

The Senate version also instructed the Department of Education initially to detail to NAGB its own staff and allowed the new organization to use up to ten percent of NAEP funds for administrative and policymaking purposes.[103]

Both the House and the Senate had passed their own versions of the ESEA reauthorization in 1987 and everyone expected a completed bill in early 1988. But there was considerable dissatisfaction among several major educational associations about the proposed expansion of NAEP. Arnold F. Fege, director of governmental relations of the National Parent-Teachers Association (PTA) stated that "enough is enough. This bandwagon of testing is getting ridiculous." And Bruce Hunter, associate executive director of the American Association of School Administrators (AASA), complained that the new plan had not been debated in the Senate and was not worth the additional $8.5 million. Hunter believed that "the marginal good to educators of comparing data across state lines, compared with the cost, is not much. The money would be better used for instruction, research, or professional development."[104]

Education Week also reported considerable disagreement between the House and Senate on the provisions relating to NAEP:

> One hotly disputed provision is the proposed expansion of the National Assessment of Educational Progress. Aides said House conferees were "apprehensive" about the Senate's NAEP proposals, and that staff members were drafting an alternative, less ambitious proposal as of late last week.
>
> The expansion plan, supported by the Education Department, calls for testing more students more frequently in more subjects, and for collecting data that allow state-by-state comparisons.
>
> The proposal for such comparisons is opposed by some educators and lawmakers, who argue that it could lead to more test-oriented instruction and result in a de facto national curriculum.
>
> One House aide said some conferees were "dead-set against" the provision and many were reluctant to "throw another $10 million" into NAEP, particularly after the recent controversy over "an anomaly" in results from the assessment's 1986 reading test.[105]

The resolution of the House-Senate differences on NAEP accepted the Senate provisions in general, but:

BEST COPY AVAILABLE

Reduced slightly the number of subjects examined.

Made the immediate use of NAEP at the state level a pilot program for the time being.

Added some additional technical expertise to NAGB.

Placed the entire operation under the supervision of the commissioner of education statistics in a newly reorganized and more independent NCES.

Although these compromises did not please everyone and would remain a source of some tension, they allowed everyone to agree in the short run and enabled the final passage of the ESEA reauthorization legislation.[106]

The final legislation, the Augustus F. Hawkins-Robert T. Stafford Elementary and Secondary School Improvement Amendments of 1988 (P.L. 100–297) passed both chambers and was signed into law in April 1988. It stated that the National Assessment would:

> collect and report data on a periodic basis, at least once every 2 years for reading and mathematics; at least once every 4 years for writing and science; and at least once every 6 years for history/geography and other subject areas selected by the Board; collect and report data every 2 years on students at ages 9, 13, and 17 and in grades 4, 8, and 12; report achievement data on a basis that ensures valid reliable trend reporting; include information on special groups.[107]

Rather than providing state-level tests in all of these subject areas, the legislation called only for trial assessments in mathematics and reading:

> The National Assessment shall develop a trial mathematics assessment survey instrument for the eighth grade and shall conduct a demonstration of the instrument in 1990 in States which wish to participate, with the purpose of determining whether such an assessment yields valid, reliable State representative data.
>
> The National Assessment shall conduct a trial mathematics assessment for the fourth and eighth grades in 1992 and...shall develop a trial reading assessment to be administered in 1992 for the fourth grade in States which wish to participate, with the purpose of determining whether such an assessment yields valid, reliable State representative data.[108]

The legislation also called for the commissioner of education statistics to contract with a nationally recognized organization such as NAS or NAE for an independent assessment of the state-level pilot programs.

In some important areas, the organization and composition of NAGB was altered from the original Senate bill. Most of these changes reflected the efforts of the House to create a more important and independent NCES and its concerns about the expansion of NAEP at the state level. Rather than having NAGB oversee the assessment contractor directly, as suggested in the Senate bill, the final legislation gave that responsibility to the new commissioner of education statistics:

> With the advice of the National Assessment Governing Board..., the Commissioner shall carry out, by grants, contracts, or cooperative agreements with qualified organizations, or consortia thereof, a National Assessment of Educational Progress. The National Assessment of Educational Progress shall be placed in the National Center for Education Statistics and shall report directly to the Commissioner for Educational Statistics.[109]

The new legislation did specify that "the National Assessment Governing Board shall formulate the policy guidelines for the National Assessment." It also listed the same set of responsibilities for the Board that had been set forth in the initial Senate bill—still giving NAGB considerable power and independence, but introducing additional potential tension between NAGB and NCES. One seemingly minor alteration in the wording, but perhaps quite an important change in the long run, was changing the call for the development of "feasible achievement goals for each age and grade" to "identifying appropriate achievement goals for each age and grade in each subject area to be tested."[110] During the debates over the advisability of setting student achievement levels in the early 1990s, NAGB emphasized what should be the "appropriate" levels rather than what might have been "feasible" to expect of students at the time.

Whether this change in the wording of the final legislation was perceived as important at that time is unclear, but that slight alteration may have assisted those who hoped to develop performance standards for students in the future.

The House generally accepted the Senate's suggestions for the procedures for selecting NAGB members as well as for the types of individuals to be appointed. The final legislation added three people to the proposed twenty-member board—another classroom teacher (so that each of the three grade levels covered by NAEP would have an appropriate teacher on the board); another curriculum specialist; and an additional testing and measurement expert.[111] Given the curriculum and testing complexities that would confront the board in the early 1990s, the addition of these three members proved to be helpful.

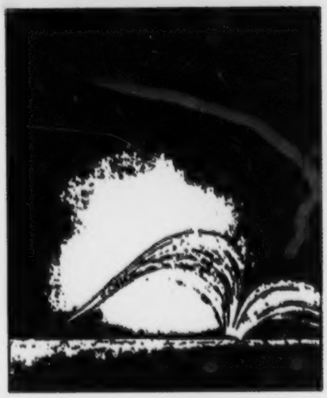
Apprehension about gathering too much inappropriate background information had surfaced in the Alexander-James report and had been mentioned in the Senate bill as well. The final legislation reiterated this concern:

> The National Assessment shall not collect any data that are not directly related to the appraisal of educational performance, achievements, and traditional demographic reporting variables, or to the fair and accurate presentation of such information.[112]

One of the major reasons why many policymakers had sought state-level NAEP data was to use the data for the Department of Education's controversial, but popular, wall charts (instead of the SAT and ACT scores, which everyone agreed were inappropriate for state comparisons). But although some people applauded these state-by-state comparisons, others strongly opposed them. Congress inserted advisory language in the final conference report that tried to prohibit the possible use of the state-level data for ranking state educational systems:

> The conferees wish to emphasize that the purpose of the expansion of NAEP is to provide policy makers with more and better state level information about the educational performance of their school children so that participating states might better measure the educational performance of their children. The goal is not to provide a scorecard by which to rank state educational systems. Data from this assessment is not to be used to compare, rank or evaluate local schools or school districts.[113]

Looking back, perhaps the most amazing fact was that almost twenty years after NAEP was created, Congress and the Reagan administration were able to come together so quickly to make fundamental changes in the operation and orientation of NAEP—especially since the Alexander-James report, which played such a key role, had been issued only a few months before both the Senate and the House finalized their particular versions of the ESEA reauthorization.



IV The NAGB Members

The law (P.L. 100–297) establishing NAGB was signed in April 1988, and called for the secretary of education to solicit nominations from various associations and organizations for members to the board.[114] All P.L. 100–297 programs were to take effect on July 1, 1988, but a technical amendment (H.R. 4638) changed the effective date to October 1 to delay the introduction of the new Chapter 1 grant formulas.[115] Secretary Bennett, who had already indicated that he was leaving the Department of Education, appointed the twenty-three members of NAGB in early September 1988—almost a month before the new law was scheduled to go into effect and just prior to his own resignation.[116]

The legislation stipulated that the members of the Assessment Policy Committee would become members of NAGB for the remainder of their current terms. The remaining slots were to be filled by the secretary from nominations by state governors, chief state school officers, education associations, parent organizations, learned societies, and NAE. Thereafter, as vacancies occurred, the board would send to the secretary the names of three individuals for each position after consulting with the groups named above.[117] The nominating procedure was amended in the 1994 reauthorization to give more direct influence to outside groups and organizations, who now could nominate up to six individuals for each vacancy in their own area of expertise and interest.[118] This was a major change in the nomination process and no longer allowed the board to perpetuate itself. In practice, however, the advice of NAGB was still quite influential. Current Secretary of Education Richard W. Riley, formerly a member of NAGB, asked the board to solicit the suggestions for the openings and then to submit "a list of six candidates for each such vacancy, who were nominated by the appropriate organization."[119] Since no clear definition existed of the type or number of groups that could nominate potential board members, NAGB retained considerable de facto power by being able to select which six nominations would be forwarded to the secretary (though this de facto power might disappear under a different secretary of education).

The 1988 legislation called for a twenty three-member board to serve for four-year terms, with no limitation on the number of times that a board member could be reappointed. The law specified the particular categories from which members of the board were to be selected:

- Two Governors, or former Governors, who shall not be members of the same political party.
- Two State legislators, who shall not be members of the same political party.

- Two chief state school officers.
- One superintendent of a local educational agency.
- One member of a state board of education.
- One member of a local board of education.
- Three classroom teachers representing the grade levels at which the National Assessment is conducted.
- One representative of business or industry.
- Two curriculum specialists.
- Two testing and measurement experts.
- One nonpublic school administrator or policymaker.
- Two school principals, one elementary and one secondary.
- Three additional members who are representatives of the general public, including parents.[120]

The contested reauthorization of NAGB six years later expanded the number of board members to 25 individuals by adding a third testing and measurement expert and a fourth representative of the general public. The legislation also attempted to promote more rotation in office by reducing the length of the appointment to three years and prohibiting members from serving more than two terms.[121] To keep one of the present board members longer, the Department of Education interpreted the law to allow the possibility of renewing the then-current NAGB members for an additional two terms.

A common but disputed perception among some Washington observers is that NAGB has remained under the control of a few particularly active members, such as Chester Finn (one of the more influential leaders behind the initial creation of NAGB as well as its first two-term chair), who have managed to preserve the policies of the initial board. NAGB has also been characterized by some policymakers as representing a generally partisan Republican belief in the need for setting high performance standards and using comparisons of state test scores to spur educational reforms.[122]

From its beginning, however, the board has tried to pursue a balanced, bipartisan orientation—partly due to the legal necessity of having equal political representation for some of the appointments but mainly because NAGB has worked hard to maintain a bipartisan stance over the years. The initial selection of the board and its chair did reflect the direct influence of Secretary Bennett and his OERI assistant secretary, Finn—though the composition of the board was also affected by the mandated presence of six holdovers from the former Assessment Policy Committee. When NAGB submitted its first set of ranked nominations to the new secretary of education, Lauro F. Cavazos, none of NAGB's first choices were appointed. In the first two years none of the existing members were reappointed.[123] Overall only 15.7 percent of the members have served more than one term.[124]

All of the appointees currently on the board were appointed or reappointed by Secretary Riley—ensuring that the Clinton administration, as was true of its predecessors, has had ample opportunity to influence the selection of the board and the general direction of NAGB's policies. Rather than being a partisan committee, NAGB is thus more of a hybrid: the product of the administration that has appointed or approved its membership, balanced by the built-in bipartisan representation it is required to have and the efforts it has made to remain bipartisan in its outlook and actions.

Although NAGB is more bipartisan and open to possible changes than some critics believe, its members have been unusually enthusiastic and consistent in their general support for NAEP and the overall policy directions of the organization. There are several possible explanations for the surprisingly stable consensus among members about NAGB's policies. Initially, most board members were probably selected because they appeared to agree with the general goals and orientation for NAEP as set forth in the Alexander-James report. When vacancies arose, either these individuals were reappointed or persons with similar views were selected. As it turned out, several subsequent secretaries of education (Lamar Alexander and Richard

Riley) continued to be strong proponents of NAGB and worked hard to ensure its continuation and success.[125] A few long-term NAGB members, such as Mary Blanton, Chester Finn, Mark Musick, and William Randall may have had a disproportionate impact, but usually as a result of their own intellectual and personal leadership rather than because of any particular political or ideological orientation. The operation of NAGB, which places heavy emphasis on involving everyone and reaching major decisions through consensus, has meant that new members are quickly familiarized with past decisions and traditions while at the same time allowed ample opportunity to influence future policies. NAGB executive director Roy E. Truby and the rest of the staff have contributed to this overall relative harmony and consensus by being attentive to the interests and ideas of the board without trying to force key policy decisions in certain predetermined directions.

Some of NAGB's overall consensus and harmony could diminish if a future secretary of education does not share the overall goals and approaches of the current group. For example, if a secretary of education was appointed who questioned the value of setting performance standards for NAEP tests, or thought that under no circumstances should NAGB be involved in developing and implementing a voluntary national test, that secretary might appoint individuals to the board who shared his or her basic orientation on these potentially divisive issues. The board might then become more divided and less able to reach a consensus. Although the current structure and culture of NAGB would help to overcome or limit these potential divisions, over time the nature of the agency and its policies could change. The Department of Education and NAGB have been careful to try to ensure that the board has representation from different regional, ethnic, gender, and political groups, and, so far, they have not appointed members with such widely differing opinions on key issues that NAGB's ability to reach near unanimous agreement has been seriously challenged.

Although more divergent elements could be included in its membership, the legislation establishing NAGB stressed the need for balanced representation:

> The secretary and the Board shall ensure at all times that the membership of the Board reflects racial, gender and cultural balance and diversity and that it exercises its independent judgment, free from inappropriate influences and special interests.[126]

Throughout NAGB's first decade of operation, the Department of Education paid close attention to this requirement and generally succeeded in achieving the desired balance and diversity in the characteristics of the members.

In terms of regional representation, most members have come from the East North Central (17.9 percent), the South (17.9 percent), the Middle Atlantic (16.4 percent), and the Pacific (16.4 percent) regions.[127] The smallest representation was from the External States and Territories (3 percent) and the Border States (6 percent). A comparison of the regional distribution of board membership with the population of those regions in 1990 shows that six of the regions are slightly overrepresented, with only the South and the Border States underrepresented (although the latter two regions had a combined total of 33.6 percent of the overall inhabitants, only 23.9 percent of the board came from either the South or the Border States).[128]

Both men and women have been active and effective NAGB members. Sixty percent of the 70 members have been males. There was little difference in the likelihood of males and females being reappointed for another term: 16.7 percent of males were reappointed to another term and 14.3 percent of the females were reappointed. Female members have often headed important NAGB committees and have been elected by the Board to serve as vice-chair; but so far all chairs have been male.

The gender of the members was relatively clear, but it was more difficult to define and interpret other characteristics, such as race and culture. Daniel B. Taylor, then the deputy executive director of NAGB, provided some useful suggestions for dealing with this problem in 1989:

> I have done a quick matrix of the current membership of the entire Board, and of those whose terms expire in 1990. I have had to make some arbitrary decisions

> in establishing the various categories of representation, and the committee might want to arrive at different definitions. About the only straight-forward category is Male/Female. For the "race and culture" categories, I have done what I think most others typically do and that is to broaden the definition of race from the generally accepted three—White, Black and Asian—to five, including Native American and Hispanic, although the latter two are not technically "races." They can, however, be considered as "cultures" and thereby satisfy the legal requirement in regard to cultural balance and diversity. Beyond Native American (which includes Alaskan native) and Hispanic, I don't think it is necessary to identify "other" cultures for purposes of representation on NAGB.[129]

NAGB annually has provided its nominating committee with information about the characteristics of the board. For example, in November 1990, the membership of the board had nineteen whites, three blacks, two Hispanics, and no Asians or Native Americans. The previous year the Secretary had "requested [that] additional names of nominees be sent to him reflecting a greater mix of racial/ethnic representation."[130] Since information about the racial and ethnic backgrounds of all board members was not readily available, an analysis of the overall racial and ethnic composition of the group during the past decade was not undertaken for this study.

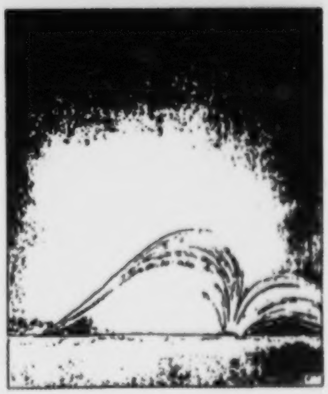
The work expectations for the board are very demanding and require extraordinary dedication and effort. The board meets at least four times a year and there are often additional committee meetings. Subgroups are expected to do additional useful but time-consuming tasks, such as reviewing all proposed questions for NAEP tests. The anticipated extra work associated with NAGB's oversight of the proposed voluntary national test will make service on the board even more onerous and may discourage some individuals from agreeing to be considered for membership.[131]

Given the heavy workload and the already busy schedules of its members, NAGB has maintained a relatively high participation rate at its meetings. Members have attended 81.9 percent of the sessions—a respectable figure for a board with so many active and distinguished members.[132] Attendance has decreased in recent years—from 85.2 percent in 1990–94 to 78.3 percent in 1995–98.[133] Whether this decrease in attendance rates reflects a reaction against a perceived increase in the workload, less interest because the major issues no longer seem as threatened and compelling, or the appointment of some members who are slightly less committed to the entire enterprise than were earlier members is not clear.[134] With the expected increased responsibilities of the proposed voluntary national examinations, it may become even more difficult to maintain the same high rates of participation in the future.

There were significant variations in the rates of participation by the different categories of members. Members who represented teachers, the general public, testing and measurement experts, and local school boards attended at rates of more than 90 percent.[135] The governors, however, attended only 21.9 percent of the time; local school superintendents 66.7 percent; and school principals 77.8 percent.[136] Members in the other six categories attended at rates of more than 80 percent.[137]

After Secretary Bennett announced the twenty-three proposed members for the board in September 1988, Augustus F. Hawkins (D-CA), chair of the House Education and Labor Committee, and thirteen other Democratic legislators declared that three of the nominees did not meet "the high standards of expertise and balance intended by Congress."[138] They cited conflicts of interest for the individuals and pointed out that the two "testing and measurement experts" lacked adequate qualifications. As they put it, "the board's testing and measurement experts should be leading figures in the field of psychometrics, should represent a diversity of approaches, and should not be closely identified with a firm conducting the NAEP." The members of Congress warned that "if the process of governing NAEP is politicized, if it becomes the plaything of those who would use federal funding to test their own pet ideas of what works, the value of NAEP will be destroyed." They then called on Cavazos, the new Secretary of Education, to withdraw the nominations of the three individuals.[139] Not only were the names not withdrawn, but Cavazos, assuring Hawkins that careful review had revealed no conflict of interest, reappointed Chester Finn as the chair of NAGB the following year; and Secretary Riley named Mark Musick as chair several years later.[140]

This episode and its aftermath exacerbated an already strained relationship between NAGB and several influential House Democrats on the Education and Labor Committee. It also reinforced, at least in the short term, the impression that the Department of Education and NAGB did not take seriously the congressional injunction that the board was to appoint experts in testing and measurement. The subsequent appointments of distinguished testing and measurement specialists such as Jason Millman (1992), Michael Nettles (1992), and Edward Haertel (1997), however, helped to diffuse these particular criticisms.



V Staffing and Financing NAGB

NAGB began operation on October 1, 1988. The first board meeting occurred on November 18-19, 1988, just seven weeks after the law went in effect, and the members quickly called for the hiring of staff and consultants to help them. They set up two working groups—the first to consider organizational and staffing needs and the second to identify key upcoming NAEP policy issues.[141]

The working group on organization and staffing met a month later and was briefed on federal procedures for hiring staff[142] by Emerson Elliott, the acting commissioner of education statistics. The working group established nine criteria for hiring an executive director—the three most important being knowledge of and involvement with the educational community, knowledge of testing issues, and consensus-building skills.[143] After a quick but thorough search, the board hired Roy E. Truby, an experienced educator and administrator who has been a public school teacher, the state school superintendent in both Idaho and West Virginia, and a visiting professor of education at the University of Arkansas at Little Rock, to be the executive director of NAGB.[144]

Daniel Taylor was hired as the deputy executive director based on his extensive educational and administrative experiences as a state school superintendent in West Virginia, a senior lecturer at the Harvard Graduate School of Education, an assistant secretary for vocational and adult education in the Department of Education, and the chief operating officer for the College Board. He remained as deputy executive director of NAGB until 1997 when he was replaced by Sharjf Shakrani, a former specialist in measurement and evaluation in the Michigan state government who had come to Washington as the chief of design and analysis at NCES.

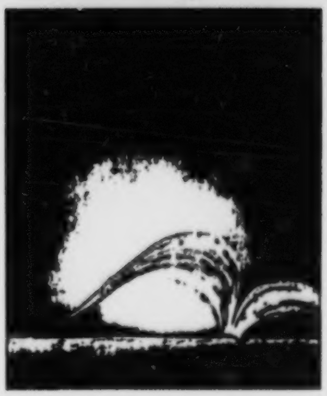
In addition to Truby and Shakrani, the NAGB staff today consists of Mary Lyn Bourque, assistant director for psychometrics; Mary Crovo, assistant director for test development; Ray Fields, assistant director for policy and research; Lawrence Feinberg, assistant director for reporting and analysis; Stephen Swearingen, budget and finance officer; and Mary Ann Wilmer, operations officer. The staff also includes two assistants, Jewel Bell and Dora Drumgold, and NAGB is in the process of hiring additional personnel to help with the proposed voluntary national test.[145] Most of the NAGB staff have been with the agency since its inception and they are well regarded by most knowledgeable outside observers—although some questions have been raised in the past about whether there were enough technical experts to handle the workload.

NAGB was assigned a significant amount of work; therefore the legislation in 1988 stipulated that funds not to exceed ten percent of the total NAEP budget could be used for administrative and policy purposes.[146] During the early 1990s, NAGB usually used approximately ten percent of the total assessment monies. As the costs of the state-by-state NAEP tests grew, so did the administrative budget of NAGB—from $938,000 in FY 1989 to $2,990,000 in FY 1992.[147]

During hearings before the House Appropriations Subcommittee, Representative David Obey (D-WI), generally a strong supporter of spending for educational statistics, expressed some dissatisfaction about the seemingly automatic increases for NAGB expenditures. Christopher Cross, the OERI assistant secretary, assured Obey that the current funds were being well spent, but he agreed that there should be a separate authorization for NAGB rather than funding based on a percentage of the overall NAEP budget.[148] When legislation was renewed in 1994, it included a separate authorization for NAGB.[149]

Some House members were dissatisfied with NAGB and quietly tried to limit the appropriations for its annual operations to $1 million for FY 1991, FY 1992, and FY 1993—a sum far less than the board believed it needed to function. The House passed the legislation in 1990 and waited for the Senate to do the same (as some Senators informally had indicated they planned to do). The Senate, however, failed to act on the legislation before adjournment. Chastened by the close call, board members then worked more closely with congressional staff to inform them about the activities and needs of NAGB to forestall any similar criticism of their administrative and policy funding.[150]

The separate authorization for NAGB did not change the NAGB appropriation much in the following years; it continued to be about 9 to 10 percent of the overall NAEP budget. Since there were no major additions to the NAEP budgets from FY 1992 to FY 1997, NAGB funding remained fairly constant during these years and saw an increase of only $606,000 in FY 1998 to cover the additional expenses associated with preparations for a possible voluntary national test. In constant dollars, however, the allocation for FY 1998 was actually slightly less than in FY 1992—even though the workload had been substantially expanded.[151] At a time when the board was trying to redesign and improve the operations of NAGB and NAEP and take on additional duties, funding for that operation remained basically the same in real dollars.



VI Interpreting NAGB's Authority and Responsibilities

One of the continuing tensions in the administration and development of NAEP is the division of labor and authority between NAGB and NCES. Although the Alexander-James study group had recommended that NAEP's oversight body be held accountable, it also wanted NAGB to be independent of the Department of Education—therefore the report called for a system of checks and balances:

> The governance and policy direction of the national assessment should be furnished by a broadly representative Educational Assessment Council that provides wisdom, stability, and continuity; that is charged with meshing the assessment needs of states and localities with those of the nation; that is accountable to the public—and to the federal government—for stewardship of this important activity; but that is itself buffered from manipulation by any individual, level of government, or special interest within the field of education.... A separate test contractor under contract with the federal government should handle test development, administration, analysis, reporting, maintenance of item banks, and provision of assistance to states and others in supplementing tests. It would be guided by policies established by the council concerning test domains, learning objectives, test design, and plans for analysis.
>
> Thus the overall governance of the nation's report card would consist of three major elements, each with specific duties, powers, and rights: the Educational Assessment Council, the testing contractor, and the federal government. This structure is meant to supply needed checks and balances and "separation of power" for this important and sensitive enterprise.[152]

The Alexander-James study group envisioned an independent governing agency and a separate assessment contractor. The governing agency "would define content areas, assessment procedures, and guidelines for fair comparisons of states and localities."[153] Although the oversight agency would be accountable to the public and the federal government, it was not envisioned to be directly controlled by any particular federal unit or individual. But the lines of authority were not entirely clear, as the test contractor was to be selected, funded, and monitored by the federal government, not by the new independent governing agency.[154]

The Alexander-James study group recognized and even welcomed the split authority, especially between the independent governing group and the testing contractor. The federal government was seen as a third party, but it was not entirely clear from the report what its overall policy or administrative roles would be or which federal agency should be involved. Nor was there a discussion of what would happen if any two or three of these participants could not reach a negotiated settlement of their differences:

> The Education Assessment Council we propose would differ from the present Assessment Policy Committee in several important respects. First, we believe that it is essential to separate the Educational Assessment Council from the test contractor. This would establish a set of checks and balances among the three entities involved in the assessment: the council, which sets testing policy and test specifications, the test contractor, who develops and administers the actual tests, and the federal government, which provides funding and awards the contract. We believe that negotiations among these three groups will strengthen the decision-making process by reflecting an array of education, measurement, and policy perspectives.[155]

The National Academy of Education, which had been commissioned to review the work of the Alexander-James study group, endorsed the idea of a strong, independent NAEP governing board, but warned about the dangers of the ambiguity in the specifications set forth in the report:

> The actual relationship of the Secretary of Education to this new council remains somewhat ambiguous in the Alexander-James report. It is not clear whether the secretary would be constrained to frame the testing contract according to the "policies and specifications" set by the EAC [Educational Assessment Council] or whether the secretary could regard these policies and specifications as merely advisory and ignore them. Since this issue remains unclear, it is only prudent to assume that the latter possibility exists. If so, the entire endeavor is left open to possible inappropriate intrusion.
>
> We recognize that a government agency cannot allow an independent organization such as the EAC to dictate to the secretary the extent of a contract for which the secretary is fiscally responsible. Beyond this fiscal control, however, the contract should follow the specifications set out by EAC. This is clearly the intent of the Alexander-James Study Group.[156]

When Congress addressed these issues in 1987–88, the two chambers were deeply divided. The House wanted a strong, independent NCES and implicitly assumed that NAEP would continue to be administered by that agency. The House did not discuss the oversight of NAEP in any great detail and certainly did not follow the recommendations of the Alexander-James study group. The Senate, on the other hand, embraced the Alexander-James report and voted to create a powerful and independent NAGB that would direct NAEP without much oversight or assistance from NCES.

The final resolution, however, left the situation much more ambiguous. The board was to "formulate the policy guidelines for the National Assessment." The law specified the board's responsibilities in some detail:

> (6)(A) In carrying out its functions under this subsection, the Board shall be responsible for—
>
> (i) selecting subject areas to be assessed (consistent with paragraph (2)(A));
>
> (ii) identifying appropriate achievement goals for each age and grade in each subject area to be tested under the National Assessment;
>
> (iii) developing assessment objectives;
>
> (iv) developing test specifications;
>
> (v) designing the methodology of the assessment;
>
> (vi) developing guidelines and standards for analysis plans and for reporting and disseminating results;
>
> (vii) developing standards and procedures for interstate, regional and national comparisons; and
>
> (viii) taking appropriate actions needed to improve the form and use of the National Assessment.[157]

At the same time, however, the legislation placed NAEP in NCES, reporting to the commissioner for education statistics. It also gave the commissioner the authority to oversee the grant or contract for carrying out NAEP:

> With the advice of the National Assessment Governing Board established by paragraph (5)(a)(i), the Commissioner shall carry out, by grants, contracts, or cooperative agreements with qualified organizations, or consortia thereof, a National Assessment of Educational Progress. The National Assessment of

BEST COPY AVAILABLE

> Educational Progress shall be placed in the National Center for Educational Statistics and shall report directly to the Commissioner for Educational Statistics.[158]

The commissioner of education statistics was also charged with providing for periodical independent evaluations of NAEP. In the original legislation, the commissioner was to "provide for an independent evaluation conducted by a nationally recognized organization (such as the National Academy of Sciences or the National Academy of Education) of the pilot programs to assess the feasibility and validity of assessments and the fairness and accuracy of the data they produce."[159] Similarly, the NAGB reauthorization in 1994 mandated that the controversial performance standards only be "used on a developmental basis until the commissioner determines, as the result of an evaluation under subsection (f), that such levels are reasonable, valid, and informative to the public."[160]

Being a legislative compromise, it is not surprising that the relationship between NAGB and NCES was not entirely clear and that no specific mechanism had been created for resolving major differences. Despite the split responsibilities and the ambiguities in the legislative language, NAGB and NCES usually have worked closely and harmoniously together in developing, implementing, and evaluating NAEP. Although there were undoubtedly some inefficiencies due to the division of labor between NAGB and NCES, there were also important benefits in having these two agencies work together, in addition to providing a check on each other in dealing with highly sensitive issues. Although some factors behind certain tensions and disputes between NAGB and NCES will be considered, the working relationship between them has been quite successful and indications suggest that further improvements are underway.

During the past decade both NAGB and NCES have been highly professional and well regarded by most observers. Yet some differences in their perspectives and orientation toward NAEP may have affected the relationship between the two organizations. NCES, a strong supporter of NAEP, seemed to be particularly concerned about the technical quality of the tests and was reluctant to release data and analyses that it believed had not been thoroughly developed and evaluated. NAGB also valued the technical validity of NAEP, but it sometimes seemed to be more willing to implement new innovations in areas such as state-level testing and setting performance standards before the instruments were fully piloted and rigorously evaluated. NAGB was also concerned that the NCES adjudication process was unnecessarily slow and contributed to long delays in providing results to policymakers and to the public.

NCES valued technical knowledge and provided ample opportunities for the experts on its staff to be included in its decision-making processes. Although NAGB also appreciated technical assistance and maintained qualified specialists on its staff, some influential board members downplayed the need for testing and measurement experts on the board itself—assuming that the necessary technical assistance and guidance could be obtained as needed. Although both NAGB and NCES recognized the necessity and importance of technical competence and assistance, some NAGB members were more willing to obtain it through special, ad hoc panels and consultants rather than by having such experts on the board.

There was also a difference between NAGB and NCES regarding the extent to which each agency thought it should be involved in doing interpretive studies for policymakers. Under the leadership of Emerson Elliott, NCES was reluctant to become too engaged in policy analyses, especially in the more controversial areas. NAGB, on the other hand, appreciated the limits of policy analysis but seemed more willing to use NAEP to further educational reforms by setting performance standards to spur student achievement and supporting more in-depth analyses of NAEP data to help policymakers improve schooling. Some differences in the willingness of NAGB and NCES to engage in analyzing data for policy-related questions are diminishing, as Pascal Forgione, the new commissioner of education statistics, has placed more emphasis on the role of NCES in the area of policy analysis and dissemination.

Another way the two units differed was that NAGB tended to focus almost all of its attention and resources on NAEP, while NCES had a broader responsibility for overseeing and developing other statistical data and analyses. As a result, NCES sometimes questioned the relative value of adding more funding for NAEP rather than spending those monies on alternative data sources.

There also may have been some differences in the leadership style of NAGB and NCES. Decisions by NAGB were made after a period of discussion, with input from all sides. Persuasion and compromise led to consensus. Decisions by NCES fell to Elliott alone to make, based on his interpretation of the work of his staff, recommendations from outside advisory groups, and his own critical analysis. Projects that required joint action by NAGB and NCES brought together these two management styles and a certain amount of friction occasionally resulted. Interactions between Elliott and NAGB were always cordial, professional, and effective, but neither may have been totally comfortable with the other's styles. Forgione, who worked closely with governing boards in his earlier posts as the director of assessment in Connecticut and then superintendent of schools in Delaware, appears to be particularly experienced and skilled in dealing with a board such as NAGB.

As NAGB initiated a redesign of NAEP in the mid-1990s, NCES cooperated by commissioning KPMG PEAT Marwick LLP to analyze the operation and management of NAEP in October 1995.[161] Specifically, NCES asked the firm and its subcontractor (Mathtech, Inc.) to focus on four tasks:

- The choice of funding (procurement) vehicle (contract, grant, or cooperative agreement) and associated management issues.
- Cost allocation and cost-tracking methods.
- Decisionmaking processes.
- Cost effectiveness and appropriateness of NAEP statistical methodologies.[162]

The report found problems in the decisionmaking process ranging from a lack of clarity in the NAEP mission statement to extra costs associated with a consensual management style:

> There appear to be three interconnected broad issues, each with negative implications on NAEP operations. This collectively creates a nearly unworkable structure, almost guaranteed to be mistake prone, high cost, slow and full of continuing controversy. It is remarkable that the participants in the enterprise do as well as they have.[163]

Yet it was the criticisms of the strained and unclear relationship between NCES and NAGB that attracted the most attention and controversy:

> The existence of the National Assessment Governing Board (NAGB) recommended in concept by the Alexander-James report and enacted into law in 1988 is to our knowledge a unique structure among federal statistical agencies. Advisory bodies with a comprehensive scope are commonplace among federal statistical agencies; multi-person groups with serious governing authority are not. NAGB's authority to oversee and give certain direction to NCES about NAEP exists in parallel with the Commissioner of NCES' authority to direct and execute the NAEP assessments....Without regard for the relative merit of the govern-ance concept, the resulting conflict between NAGB and NCES over authority to decide each issue has substantial and largely unmeasurable consequences for NAEP in time and cost. This confusion about executive decision authority is made far worse by the ab-sence of any established dispute resolution machinery which prolongs the duration of any disagreement and can result in repeated conflicts over the same issue on subsequent occasions.
>
> The lack of dispute resolution mechanism between NAGB and NCES encourages a further management structure problem, which we observed though not often. That problem is the involvement of the senior political levels of the Department of Education in the decisions about the substantive content of assessments.[164]

The Marwick and Mathtech report offered several recommendations for dealing with the decisionmaking problems it detected:

> The selection of which subjects to assess (and how often) and the determination of performance standards, if continued, are matters of high policy content in which the NAEP stakeholders have critical interests. There is, therefore, in our view a case to be made for a body external to the agency with executive authority for NAEP as a whole to have decision-making authority with respect to these particular matters. These are functions which NAGB has and could, presumably, continue to perform.
>
> The NAGB-NCES interface with respect to all other matters needs to be clarified. In business terms, NAGB has been seeking to function as the CEO rather than the Board of Directors of the NAEP enterprise. The advice and counsel of the stakeholders remains important across a wide range of these other matters, but it should be advice and counsel not decision-making, more like the model of other large scale information gathering projects.... A further step toward the improved operations and the resolution of the full range of management structure problems would include raising the visibility of the Commissioner of NCES as the key senior official with respect to NAEP including the resolver of disputes and protector of its integrity and nonpartisan character. As a related matter, it is important to the basic viability of NAEP that neither NAGB nor senior Departmental officials be seen as possible or effective interveners on behalf of specialized ideological agendas or partisan interests.[165]

Overall, NCES was pleased with the Marwick and Mathtech report. Acting Commissioner of Education Statistics Jeanne Griffith praised the analysts: "I think they provide a lot of insight for change and improvement for all pertinent parties who are all together trying to achieve a more efficient and effective and useful NAEP."[166] And Forgione, the current commissioner, accepted the Marwick and Mathtech recommendation on how to improve the decisionmaking problems:

> *Recommendation Accepted*: NCES believes that this recommendation can be addressed by revisiting the NAEP legislation and specifically delineating its operational implications for major institutional actors. All significant parties, including the Commissioner of Education Statistics, the Advisory Council on Education Statistics (ACES), NAGB, senior level Department officials, and their respective staffs, need to have a shared understanding of the management structure currently defined in the law. We agree, for example, that the NAGB-NCES interface needs to be clarified and that conflict resolution resides with the Commissioner. To review the legislation critically in this regard, we believe, would be the first step towards improved operations.[167]

NAGB strongly disagreed with some of the basic premises of the Marwick and Mathtech report. NAGB agreed that management and decisionmaking problems existed and needed to be addressed, but it believed the analysts had failed to consider the historical and legal policy rationale for the organization. William Randall, chair of the board, responded to a draft of the Marwick and Mathtech report:

> The draft report of Peat Marwick's NAEP Management Review has two serious shortcomings in regard to the National Assessment Governing Board:
>
> > 1. It shows little understanding of the policy rationale for establishing the Governing Board and of the extensive responsibilities given it by law (P.L. 103–382).
> >
> > 2. It virtually ignores the Board's major strategic planning initiative, begun in November 1994, which has produced a plan for redesigning NAEP to sharpen its focus and simplify its design to enable the assessment to test more subjects more frequently, release reports more quickly, and reduce costs. Over the next few months the Board is making extensive efforts to solicit public comment and expert review before taking final action on the redesign policy at its August meeting.
>
> Because of these shortcomings the draft report often misconstrues the proper relationship between NAGB and the National Center for Education Statistics—policy formulation by the independent Governing Board and program administration by NCES, which operates as part of the Department of Education.[168]

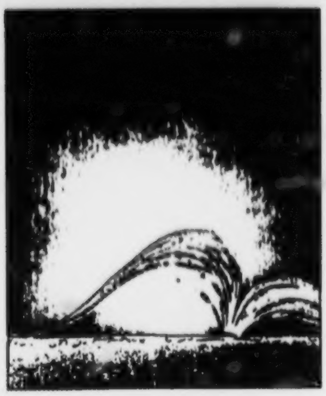
When the final Marwick and Mathtech report was released several weeks later, Mary Blanton, vice chair of NAGB, expressed her disappointment with the work and believed it had not considered the role Congress had intended for the board.

Instead of enhancing the role of the commissioner of education and NCES, Blanton suggested that it would be more appropriate to give more power to NAGB. "The fact there is this nonfederal, nonbureaucratic board that is overseeing [NAEP] gives it a lot more credence with the American people, with state testing people, with all of those people that Congress intended to be the audience for the test."[169]

Rather than focusing almost exclusively on the possibility of expanding the role of NCES to improve the management of NAEP, a recent Congressional Research Service report suggested another alternative that should be considered as well—expanding the role of NAGB even further:

> While the authors of the KPMG Peat Marwick/Mathtech study recommended resolving confusion about the authority over NAEP of NAGB versus NCES by more clearly circumscribing and narrowing the scope of NAGB's authority, an alternative option would be to increase NAGB's independence and give it authority over all NAEP policies, and possibly operational authority over NAEP as well, along with greater autonomy as a bureaucratic entity.[170]

The Congressional Research Service report then discussed several options for changing the management structure and operations of NCES and NAGB in overseeing NAEP.[171]



VII Implementing State-Level NAEP

Much of the hostility toward collecting and reporting state-level student achievement data had disappeared by the late 1980s. There was a growing interest among state governors in the collection of state student achievement information and a realization among reformers that having only national or regional data was unlikely to stimulate major educational improvements. The Alexander-James report emphasized the need for state-level results, and although the National Academy of Education (NAE) wondered about the educational impact of reporting state-level data, it did not oppose state assessments in principle.

The U.S. House of Representatives opposed the idea of authorizing state-level student achievement data, but the Senate remained steadfast in its support for assembling such information. The legislative compromise enacted in 1988 allowed NAGB to proceed with two Trial State Assessments (TSAs) of public school students. Eighth graders were to be tested in mathematics in 1990. Fourth and eighth graders were to be tested in mathematics and fourth graders examined in reading in 1992.[172]

The passage of legislation for the two TSAs did not end the debate about the value of state-level NAEP testing. Rather than awaiting the results from the congressionally mandated evaluation of the initial TSA, NAGB called for expanding state-level assessments and repealing the legislative prohibition against district testing:

> 1. *The National Assessment of Educational Progress should provide information for an annual report card by testing at least three subjects each year.*
>
> 2. *NAEP should move as quickly as feasible to full state participation in all subjects and all three grade levels (4th, 8th, and 12th) tested. No state, however, should be compelled to participate. The federal government should pay the full cost of the state-by-state NAEP program.*
>
> 3. *The Governing Board urges Congress to remove the prohibition against the use of NAEP tests and data reporting below the state level.*[173]

Opponents of an expanded NAEP believed NAGB had ignored the spirit and intent of the congressional compromise by moving forward so quickly—before the first state-level assessments had even been field tested. Paul G. LeMahieu, the immediate past president of the National Association of Test Directors, opposed the proposed expansion of NAEP and withdrew the Pittsburgh schools from participation in the 1990 pilot mathematics assessment.[174] The 389 delegates to the International Reading Association (IRA) also voted unanimously against "the proliferation of school-by-school, district-by-district, state-by-state, and province-by-province comparison assessments."[175]

The testing community remained divided on the issue of supporting state-level NAEP. Gary W. Phillips, then acting associate commissioner of the education assessment division at NCES, praised the benefits of state-by-state comparisons:

> The first important benefit of the NAEP Trial State Assessment is the information *system* it will provide. For the first time in history, we will have a reliable and valid state comparison of what students have learned in school. Not only can we compare states, but over time we can monitor state progress.... In addition to comparing states and monitoring their progress over time, we will also obtain information on whether states are doing well enough....
>
> This leads me to the second and most important benefit of the Trial State Assessment. A better information system and sustained public interest will ultimately result in improved learning for our nation's school children. In addition to finding out how well our students are learning, the Trial State Assessment will give state-by-state comparisons on the home learning environment (homework, television watching, access to reading materials), instructional practices, time spent studying, teacher and principal training and experience, educational resources and materials, composition of the student population, and demographic characteristics of the schools.[176]

On the other hand, Daniel M. Koretz, a senior social scientist at RAND, saw state-level NAEP tests as an "ill-conceived policy." Koretz pointed out the limitations in state-level descriptive data and doubted whether any reliable, causal inferences could be made about which factors accounted for the differences in state student achievement scores:

> NAEP is purely cross-sectional, which eliminates a large number of the designs that could be used to draw causal inferences. Moreover, the cross-sectional nature of NAEP means that even when differences in scores do reflect differences in programs, we won't be able to ascertain which differences in policy or practice are responsible for differences in NAEP scores. A state that has a lousy middle-school mathematics curriculum, for example, may have a strong enough elementary curriculum to score better than its neighbor on the Grade-8 state NAEP nonetheless.
>
> The NAEP also does not provide the type of data that would be required for reasonable cross-sectional causal modeling. It does not allow one to rule out other entirely plausible explanations of state differences. One reason is its limited, individual-level background data. Some important variables—such as family income—are entirely lacking. Other important variables are measured solely by student self-reports, which are known to be quite error prone even at grades higher than the eighth.[177]

Koretz argued that not only would state NAEP results yield much less useful information than its proponents believed, but that the financial costs of gathering the data were extraordinarily high. He wondered whether those monies might not be better spent on improving the national NAEP test or supporting other school improvement research projects. Koretz also feared that, if the new NAEP tests were used to hold states accountable, there would be pressure to teach to the assessments—thereby inadvertently undermining the validity of this valuable national assessment.[178]

Even as doubts continued to be expressed about the wisdom or utility of state NAEP exams, larger concerns about reforming American education arose in the late 1980s and early 1990s and reinforced the call for developing reliable and comparative state tests. The National Governors' Association (NGA) and President George Bush met at the historic education summit in Charlottesville, Virginia, on September 27–28, 1989. At the end of the session, they issued a joint statement that called for measures of progress at the level of the individual, the school, and the states:

> As elected chief executives, we expect to be held accountable for progress in meeting the new national goals and we expect to hold others accountable as well. When goals are set and strategies for achieving them are adopted, we must establish clear measures of performance and then issue annual Report Cards on the progress of students, schools, the states, and the Federal Government.[179]

The National Education Goals Panel (NEGP) was created to assemble and report data on the nation's progress toward meeting the six national goals. Given the pressing need to assemble comparable measures of student achievement, NGA passed a resolution in February 1991 calling for an expansion of NAEP to permit state-by-state and even district-by-district comparisons.[180] NEGP has relied heavily on the state NAEP results for its annual reports and has thereby lent considerable support to the collection of such information.[181]

Members of Congress also responded with increased calls for assessing progress toward the national education goals. Senator Jeff Bingaman (D-NM) introduced a bill to create a council on education goals that would monitor student progress—*The National Report Card Act of 1990*. Although it was cosponsored by Senate majority leader George J. Mitchell (D-ME) and Senator Edward M. Kennedy (D-MA), chair of the Labor and Human Resources Committee, the bill was not enacted.[182] Instead, Congress created the temporary National Council on Education Standards and Testing (NCEST) in 1991 to advise on the feasibility and desirability of national standards and tests.[183]

NCEST recommended standards and high-stakes tests for students as well as standards for schools and school systems:

> The Council concludes that the United States, with appropriate safeguards, should initiate the development of a voluntary system of assessments linked to high national standards. These standards should be created as expeditiously as possible by a wide array of developers and be made available for adoption by states and localities. The Council finds that the assessments eventually could be used for such high-stakes purposes for students as high school graduation, college admission, continuing education, and certification for employment. Assessments could also be used by states and localities as the basis for system accountability.[184]

NCEST specifically singled out the importance of NAEP and its role in helping states monitor their progress:

> The Council recommends that the National Assessment of Educational Progress (NAEP) be reauthorized and assured funding to monitor the Nation's and states' progress toward Goals 3 and 4 of the National Education Goals. NAEP is the national program begun in 1969 to biannually test representative samples of students in grades 4, 8, and 12 in core subject areas and report achievement trends over time. As the national standards are developed, there should be efforts to ensure that NAEP will be aligned with these standards.[185]

Not everyone was pleased with NCEST's strong recommendations for national, high-stakes testing. A group of prominent educators and researchers, including a few who had initially endorsed the NCEST report, rejected the recommendations and cautioned against any national or state-level tests that would hold students or their school districts more accountable.[186] And although many, if not most, policymakers expressed increased support for national and state assessment tests, some education researchers continued to question the emphasis on national content standards and aligned assessments. Linda Darling-Hammond, in an essay on the national standards, summarized her opposition to standards-based reform:

> This article argues that content standards aligned with tests are the wrong starting point for systemic school change aimed at improving teaching and learning for all students, and that national standards and assessments are the wrong vehicle. There are three reasons for this. First, top-down specifications of content linked to tests cannot take into account the many pathways to learning that will be appropriate for different students in schools across the country.... Second, national standards and tests are inappropriate vehicles for enhancing teaching and stimulating school change.... And, finally, content and performance standards are already proving themselves, once again, to be a weak, ineffectual means for leveraging resource equalization. Inequalities of learning opportunities must be addressed head-on if they are ever to be successfully removed.[187]

The National Academy of Education (NAE) was commissioned to evaluate the two TSAs. After reviewing the 1990 TSA, the NAE panel advised "that Congress should approve the continuation of state NAEP, but before legislating a permanent state NAEP, should authorize additional trials."[188] The panel rejected the idea of reporting results at the district, school, or student levels and called for private school students to be tested as well as those who have dropped out of school.[189] The NAE panel reaffirmed the same basic recommendations when it reviewed the 1992 TSA—adding more weight to the arguments of those who believed that the trial state NAEP should be extended.[190]

NAGB welcomed the support for state NAEP, but it challenged the panel's opposition to district-level assessments. Richard A. Boyd, chair of NAGB, commented:

> The Board affirms the fundamental purpose of NAEP as a monitor of student achievement, administered to national and state representative samples of students in grades four, eight, and twelve. However, the Board continues to believe that, at local option and cost and with appropriate procedures for test security and administration, states and school districts should be permitted to augment the NAEP sample and report results below the state level.
>
> The Governing Board believes, as does the Panel, that NAEP's value as an indicator of education performance should not be compromised. However, the Governing Board is unaware of direct evidence that lifting the prohibition would compromise NAEP. In fact, prior to the 1988 prohibition, reporting below the state level occurred periodically at local option and cost, and with no known erosion of NAEP's integrity.
>
> On the basis of what is known about NAEP, a recommendation against lifting the prohibition seems more a policy preference than a judgment based on data. It is well known that the reason for instituting the prohibition in 1988 had to do with political fears of federal encroachment on local autonomy associated with the advent of state-level assessment, not technical concerns about potential damage to the integrity of NAEP.[191]

Given the strong support for state NAEP as a result of the 1990 and 1992 TSAs and the endorsements of NEGP and NCEST, the Senate voted to extend the math and reading tests to the fourth, eighth, and twelfth grades in 1994 on April 21, 1993 (S. 801).[192] The House, which had been the most hostile to the state NAEP, unanimously passed the same legislation on May 11, 1993.[193] When NAGB and NAEP were reauthorized in 1994, the legislation permitted state testing at all three grade levels for all subjects, but labeled them "developmental" until the "Commissioner determines, as a result of an evaluation required by subsection (f), that such assessment produces high quality data that are valid and reliable."[194]

Although the congressional reauthorization permitted the expansion of TSAs in math and reading to all grades in 1994, budget limitations forced a dramatic reduction in the state assessments. Congress approved less than half of the $65 million requested for NAEP—approximately the same amount as it had allocated the previous year. As a result, NCES decided to cut back the state NAEP to a single test. NAGB appealed that decision to Secretary Riley and argued that a reduction in state NAEP tests "would be a setback not only for NAEP itself but also for officials and the public in many states that have begun to rely on NAEP data as an independent, valid, and comparable measure of education results." NAGB proposed reducing some of the analyses and evaluations, hoped that the contractor could reduce its own costs, and appealed for monies from elsewhere in the Department of Education to field both fourth-grade and twelfth-grade state reading tests. NCES prevailed, however, and only a fourth-grade state reading assessment was funded for 1994.[195]

The NAE panel evaluated the 1994 state assessment and praised its content validity, sampling, and assessment administration—although it did raise questions about the sample size and participation rates for nonpublic schools. The panel also encouraged NAEP to continue its efforts to include more students with disabilities or limited English proficiency; called for the reconsideration of the performance standards; and acknowledged the value of state-level data to educators and policymakers. In earlier reports, the panel had simply concluded that the state NAEP did not have a "deleterious effect on national NAEP." Now it believed that the state NAEP was actually beneficial for the national NAEP:

> [T]he Panel believes than an implicit, mostly unspoken *quid pro quo* has been developed between the states and NAGB, by means of which the states are willing to participate in national NAEP *at least in part*

> because of the value they get from participation in State NAEP.[196]

The NAE panel then endorsed the state NAEP and called for Congress to reauthorize it on a permanent basis:

> Based on its evaluation of the TSAs, the Panel concludes that state NAEP has been shown to be a valid, reliable, and useful measure of student achievement, and that it aligns favorably with the Panel's quality, utility, and state indicator principles. For these reasons, the Panel recommends that state NAEP be continued, and that it be moved from developmental to permanent status when NAEP is next reauthorized. However, in light of its size and cost, the Panel further recommends that the scope and function of state NAEP be reviewed regularly, and particularly after any substantial change in mission or design. Such re-evaluation should be done in the context of the overall NAEP program and with the abiding aim of providing the best and most useful information about student achievement for the nation.[197]

Faced with few prospects for additional funding while there was a growing demand for more state NAEP assessments, NAGB created a work group on planning in November 1994 that began to explore a redesign of the entire program.[198] The work group commissioned several analyses, drafted possible mission statements, and met several times in preparation for major discussions of NAEP's future by the entire board in 1995 and 1996. There was a general feeling that the design and administration of NAEP had grown by increments and had become unwieldy and inefficient over time. The board unanimously adopted a "Policy Statement on Redesigning the National Assessment of Educational Progress," on August 2, 1996, which candidly acknowledged the current problems:

> While there is much about the National Assessment that is working well, there is a problem. Under its current design, the National Assessment tests too few subjects, too infrequently, and reports achievement results too late—as much as 18 to 24 months after testing. Testing occurs every other year. During the 1990s, only reading and mathematics will be tested more than once using up-to-date tests and performance standards. Six subjects will be tested only once and two subjects not at all during the 1990s....
>
> The current National Assessment design is overburdened, inefficient, and redundant. It is unable to provide the frequent, timely reports on student achievement the American public needs. The challenge is to supply more information, more quickly, with the funding available.[199]

To overcome these limitations, the board set forth its goals for the future:

> The National Assessment shall be conducted annually, two or three subjects per year, in order to cover all required subjects at least twice a decade. The National Assessment shall assess all subjects listed in the third National Educational Goal—reading, writing, mathematics, science, history, geography, civics, the arts, foreign language and economics—according to a publicly released schedule adopted by the National Assessment Governing Board, covering eight to ten years, with reading, writing, mathematics, and science tested more frequently than the other subjects.
>
> The National Assessment Governing Board shall consult with technical experts and with education policymakers, in conjunction with the development of assessment frameworks, to determine the feasibility, desirability, and costs of combining several related subjects into a single assessment.[200]

The policy document specifically addressed ways to improve the state NAEP:

> National Assessment state-level assessments shall be conducted on a reliable, predictable schedule according to an eight to ten year plan adopted by the National Assessment Governing Board. Reading, writing, mathematics, and science at grades 4 and 8 shall be given priority for National Assessment state-level assessments.
>
> States shall have the option to use National Assessment tests in other subjects and at grade 12 by assuming a larger share of the costs and adhering to requirements that protect the integrity of the National Assessment program. However, the National Assessment Governing Board shall seek ways to make such use of National Assessment tests attractive and financially feasible. Where possible, changes in national and state sampling procedures shall be made that will reduce [the] burden on states, increase efficiency, and save costs.[201]

BEST COPY AVAILABLE

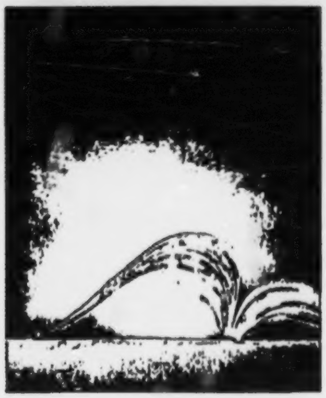
NAGB also pledged to work with states and others to link their assessments with NAEP. It also promised to help them use NAEP to improve state and local education:

> The National Assessment shall develop policies, practices, and procedures that assist states, school districts, and others who want to do so at their own cost to link their test results to the National Assessment. The National Assessment shall be designed so that others may access and use National Assessment test frameworks, specifications, scoring guides, results, questions, achievement levels, and background data. The National Assessment shall employ safeguards to protect the integrity of the National Assessment program, prevent misuse of data, and ensure the privacy of individual test takers.[202]

NAGB also spelled out what NAEP would not try to do, thereby indicating some of the limitations inherent in the types of data that were being collected:

> The National Assessment is intended to describe how well students are performing, but not to explain why. The National Assessment only provides group results; it is not an individual student test. The National Assessment tests academic subjects and does not collect information on individual students' personal values or attitudes. Each National Assessment test is developed through a national consensus process. This national consensus process takes into account education practices, the results of education research, and changes in the curricula. However, the National Assessment is independent of any particular curriculum and does not promote specific ideas, ideologies, or teaching techniques. Nor is the National Assessment an appropriate means, by itself, for improving instruction in individual classrooms, evaluating the effects of specific teaching practices, or determining whether particular approaches to curricula are working.[203]

On March 8, 1997, NAGB adopted the schedule for the national and state tests through the year 2010. The plan indicated the years for the regular national NAEP exams as well as the years in which more comprehensive assessments would occur. Every other year there would be state NAEP exams for grades four and eight, alternating between reading/writing and mathematics/science (starting with reading/writing in 1998).[204] By the end of NAGB's first decade, a permanent and stable pattern of state NAEP tests in reading, writing, mathematics, and science had been announced. The plan was accepted and almost no protests were registered.



VIII Developing NAEP Performance Standards

Disagreements over using NAEP for state-by-state comparisons have divided educators and policymakers since the late 1960s, but that issue became less contentious during the late 1980s. NAGB's unanimous decision in May 1990 to establish NAEP performance standards, however, created a great deal of controversy that seemed at times to threaten the board's survival.[205] Because no standard procedures or methods exist for setting achievement levels, differences of opinion on what those performance standards should be were not easily reconciled.[206] Debates over performance standards involved complex policy issues and often hard-to-understand technical matters. Many policymakers found it difficult to follow these arguments closely—particularly when both the proponents and opponents of NAEP performance standards sometimes appeared to focus more on defeating their challengers than resolving their conceptual and technical differences.

Setting student performance standards had not been part of NAEP assessments in the 1970s and 1980s. Some states had established minimum student competency standards in the mid-1970s, but these efforts were not tied to NAEP. Nor did the reauthorization of NAEP in 1978 provide any evidence that Congress wanted the development of performance standards. That legislation simply stated that "the Assessment Policy Committee...shall be responsible for the design of the National Assessment, including the selection of the learning areas to be assessed, the development and selection of goal statements and assessment items...."[207]

The Alexander-James study group, which proposed major changes in NAEP, hinted at but did not emphasize the need to establish performance standards. The report did include a very brief recommendation for establishing "feasible achievement goals:"

> The chief responsibility of the new council would be to shape each assessment, selecting the content areas to be tested, defining conceptually the ground to be covered in each area, setting test specifications, and identifying feasible achievement goals for each of the age and grade levels to be tested.[208]

The NAE panel, in its comments on the Alexander-James report, dealt with this subject at much greater length and did call for the development of student performance levels:

> We recommend that, to the maximal extent technically feasible, NAEP use descriptive classifications as its principal reporting scheme in future assessments. For each content area NAEP should articulate clear descriptions of performance levels, descriptions that might be analogous to such craft rankings as novice, journeyman,

> highly competent, and expert. Descriptions of this kind would be extremely useful to educators, parents, legislators, and an informed public.
>
> As NAEP continues to embody new technical advances in measurement theory, there is a real danger of getting lost in the numbers. For example, the major headings employed in the literacy report are scale score categories ranging from 150 to 400 in increments of 50 to 75. These numbers are arbitrary from both a substantive and technical point of view. Any range of values could have been employed. There is a danger of misuse of numbers like these by well-meaning policymakers who have little or no sense of their limitations.
>
> A great deal of test data is so difficult to interpret. What does a level 400 on a reading test mean? Such scores can be used for comparison across time and localities, but the nation's report card would be more broadly informative if it provided clear descriptions of the levels of competence demonstrated by our children. Much more important than scale scores is the reporting of the proportions of individuals in various categories of mastery at specific ages. In several fields, particularly reading and mathematics, we are in a position to describe beginning, average, and advanced competence at various ages. In other areas, such as writing, science, and computer literacy, research remains to be done. NAEP efforts in this area can profit both from the current endeavors of subject-matter specialists and from scientific advances in understanding student learning and cognitive skills. NAEP has already made progress in this direction, and we encourage further effort.[209]

When NAEP was reauthorized in 1988, the Senate and the House disagreed on the advisability of developing performance standards. The Senate bill (S. 373) accepted the recommendations of the Alexander-James report and called for NAGB to "identify feasible achievement goals for each age and grade in each subject area under the National Assessment."[210] The House version of that legislation (H.R. 5), however, was silent on this issue.[211] The final law (P.L. 100-297) kept the recommendation to identify achievement goals, but substituted the word "appropriate" for "feasible." The board was to be responsible for:

> [I]dentifying appropriate achievement goals for each age and grade in each subject area to be tested under the National Assessment.[212]

The phrase "identifying appropriate achievement goals" had been deliberately ambiguously worded by the Congress. Terry Hartle, the chief education staff advisor to Senator Kennedy, had played a key role in drafting the NAGB provisions in the 1988 legislation.[213] In an appearance before NAGB in May 1989, he explained that, although some people hoped that an agreement might be reached on what students should know, Congress was "deliberately ambiguous" because neither the congressional staff nor the education experts could agree on how to formulate this objective. In answering a question from a board member about the philosophical assumptions behind this directive and the concerns that this provision might lead to a federal curriculum, Hartle replied:

> The assumption there was, unless you had some ideas [of] what kids ought to know or had some idea [of] what reasonable goals would be for students...[,] it would be very hard to develop tests that could determine whether or not you were researching those goals, that somewhere there ought to be some effort to specify what we want kids to know, what we think kids should know in terms of age and grade levels. It was simply as simple as that. There was not an enormous amount of introspection on that. The concern about a federal curriculum didn't really come up very much; very infrequently did someone say: "Hey is this going to be [the basis for a federal curriculum]?" It was simply an effort to say we need to know what we're shooting for.[214]

NAGB's responsibility for "identifying appropriate achievement goals" did not immediately attract much attention. Neither the House nor the Senate conference reports on the legislation discussed the matter,[215] and few educators or policymakers at first paid much attention to the call for setting achievement goals. One notable exception was Harold Howe, II, the former U.S. Commissioner of Education and a long-time supporter of state NAEP, who worried that the ambiguous wording might be interpreted to mean that NAGB could set student achievement levels—which he did not favor or think Congress had really intended. In a letter to Emerson Elliott in May 1988, Howe warned that NAGB might use this requirement to tell schools "what their curriculum should be and what is an acceptable level of student performance in that curriculum:"

> The NAEP was created to be a service to tell Americans what young people know and can do in certain important areas of learning and how it is changing. The main objective of the new legislation was to extend that purpose to encourage state level use of NAEP. Those of us who recently supported the new legislation and its funding (myself among them) had no intention of creating a new authority to tell all American schools what to teach in each grade or even that schools should be organized by grades. More importantly, most educators are aware that any group of children of a particular age or grade will vary widely in their learning for a whole host of reasons. To suggest that there are particular learnings or skill levels that should be developed to certain defined points by a particular age or grade is like saying all 9th graders should score at or above the 9th grade level on a standardized test. It defies reality.[216]

Howe suggested that a technical amendment be passed to prevent any misinterpretation of that passage and forwarded a copy of his letter to John F. Jennings, counsel to the House Committee on Education and Labor.

While some educators and policymakers warned against setting NAEP performance standards, others encouraged NAGB to develop student achievement goals. Many states were creating their own student tests and wanted outside assistance in setting reasonable and comparable student achievement standards.[217] Influential members of Congress such as Senator Jeffrey Bingaman (D-NM) called on NAGB to help develop student performance standards.[218]

President Bush and the nation's governors met at the Education Summit in Charlottesville, Virginia, on September 27–28, 1989, and called for the establishment of student performance measures.[219] NEGP was then created and looked to NAGB for assistance in measuring student achievement.[220] In March 1990 the National Governors' Association (NGA) issued a lengthy statement on the National Education Goals in which NGA encouraged NAGB to develop performance standards:

> National education goals will be meaningless unless progress toward meeting them is measured accurately and adequately, and reported to the American people. Doing a good job of assessment and reporting requires the resolution of three issues.
>
> First, what students need to know must be defined....
>
> Second, when it is clear what students need to know, it must be determined whether they know it.... The governors urge the National Assessment Governing Board to begin work to set national performance goals in the subject areas in which NAEP will be administered. This does not mean establishing standards for individual competence; rather, it requires setting targets for increases in the percentage of students performing at the higher levels of the NAEP scales.
>
> Third, measurements must be accurate, comparable, appropriate, and constructive.... The President and the governors agree that while we do not need a new data-gathering agency, we do need a bipartisan group to oversee the process of determining and developing appropriate measurements and reporting on the progress toward meeting the goals. This process should stay in existence until at least the year 2000 so that we assure 10 full years of effort toward meeting the goals.[221]

While these other groups were discussing the need for achievement standards, NAGB was exploring how to set performance goals—and had done so from its beginning. At its second meeting in January 1989, Governor Richard Riley (SC) sought to clarify the specific responsibility of the board on the issue of performance standards:

> Are we just...involved in the technical involvement of deciding what the child does know; or do we go beyond that scope[?] Should we be into what a child should know and then develop testing mechanisms to determine if that child is learning what they should[?][222]

Chester Finn, chair of the board, replied:

> Let me try and answer this way.... We have a statutory responsibility that is the biggest thing ahead of us, to—it says here: "identify appropriate achievement goals for each age and grade in each subject area to be tested."... It is in our assignment. We have not as a Board decided how to do that.[223]

Members responded positively to these remarks and some recalled an initial discussion about setting student achievement levels during the final lunch session at the first meeting. NAGB member Saul Cooperman then remarked that "I think our job is to aspire to what it ought to be." NAGB member Herbert Walberg joined in the discussion and agreed "that we must have *ought* as well as *is*."[224]

With the increased interest in the development of student performance levels and the recognition that Congress had mandated them, NAGB moved quickly to create appropriate standards for the forthcoming 1990 NAEP mathematics assessment. It commissioned Joe Nathan, a senior fellow at the Hubert H. Humphrey Institute of Public Affairs at the University of Minnesota, to review previous attempts to set standards for student outcomes and to consider alternative ways in which the board might interpret the statute.[225] Roy Truby then produced a "Staff Paper on Setting Goals for the National Assessment" for the December 1989 NAGB meeting, in which he outlined several possible courses of action. But Truby urged postponement of a final decision to proceed with the development of performance standards until the March 1990 meeting to allow time for the staff to solicit additional outside comments and suggestions.[226]

On January 25, 1990, NAGB listened to seven hours of testimony on the proposal to create grade-level achievement goals. Assistant Secretary of Education Cross read a letter from Secretary Cavazos that endorsed NAGB's setting of performance standards. Identifying appropriate achievement goals "would be a clear definition of what constitutes grade level performance in each subject so that future (NAEP) reports could provide data on the proportion of students who achieve that standard and in what ways American students exceed or fall short."[227] Albert Shanker, president of the American Federation of Teachers, endorsed the idea of performance standards but argued against having a single standard that might cause schools to focus only on helping those near the cutoff point, while ignoring those well above or well below that mark.[228] Keith Geiger, president of the National Education Association, believed it was premature to set standards before the first state-by-state data had been collected and evaluated. He also warned about the possible dangers of "a nationally mandated syllabus."[229] And Gordon Cawelti, executive director of the Association for Supervision and Curriculum Development, believed it was more important "to get the curriculum in order before we set a decent standard."[230]

NAGB's staff drafted a series of responses to questions in preparation for the February 1990 meeting—all of which in essence were adopted at the end of the joint meeting of the committees on technical methodology/analysis, reporting, and dissemination (though not necessarily using the exact wording suggested by the staff). These detailed answers provide useful insights into the thinking at NAGB at an early stage of performance standards development. For example, the staff clarified how they interpreted the phrase "appropriate achievement" in the legislation:

> By law NAGB is required to identify goals of "appropriate achievement." Here the word appropriate is very important. Ultimately, appropriateness is a matter of taste. In its goal-setting plan NAGB intends to base its definition of "appropriate achievement goals" on knowledge and skills a consensus of educators and others say is needed to achieve the next level of subject-matter mastery. For 12th grade the Board intends to expand this consensus-building process to include employers and members of the public, college professors and scholars, to define the knowledge and skills all students need to participate in our competitive economy. We also propose to define the levels of proficiency needed to handle college-level work.[231]

Originally the staff had recommended a single standard. On the basis of outside comments as well as their own rethinking of the issue, they recommended three levels:

> In the final analysis, a single "universal" standard was recommended partially because staff believed that more than one level could be misread as "tracking" of students, [and] the Board discussion in Austin reconfirmed this belief. The testimony we received in this regard, however, was illuminating. There were persuasive arguments for several levels which would show the distribution of students and a concern that a single standard could end up as a "minimal standard."

> Staff now recommends three levels. We have been convinced that we need to set goals in such a way that it will underscore the reality of what we already know, that a distribution of performance exists and that there are enormous gaps in performance. Thus, we ought to be setting targets for the entire distribution of student performance.[232]

At the May 11, 1990, meeting, the board voted to establish three achievement levels for each grade and subject, report the proportion of students at each level, and illustrate the responses by sample items. Generic definitions for the three levels were provided:

> *Proficient.* This central level represents solid academic performance for each grade tested—4, 8, and 12. It will reflect a consensus that students reaching this level have demonstrated competency over challenging subject matter and are well prepared for the next level of schooling. At grade 12 the proficient level will encompass a body of subject-matter knowledge and analytical skills, of cultural literacy and insight, that all high school graduates should have for democratic citizenship, responsible adulthood, and productive work.
>
> *Advanced.* This higher level signifies superior performance beyond grade-level mastery at grades 4, 8, and 12. For the 12th grade the advanced level will show readiness for rigorous college courses, advanced technical training, or employment requiring advanced academic achievement. As data become available, it may be based in part on international comparisons of academic achievement and may also be related to Advanced Placement and other college placement exams.
>
> *Basic.* This level, below proficient, denotes partial mastery of knowledge and skills that are fundamental for proficient work at each grade—4, 8, and 12. For 12th grade this will be higher than minimum competency skills (which normally are taught in elementary and junior high schools) and will cover significant elements of standard high school-level work.[233]

Finn praised the board's decision to establish performance standards: "NAEP will report, for the first time in history, how good is good enough. What has been a descriptive process will become a normative process."[234] But Paul G. LeMahieu, director of the Division of Research, Evaluation, and Test Development for the Pittsburgh Public Schools, disagreed: "NAGB ought to be about the business of...ensuring the quality and innovative character of NAEP. They're distracted from that in their quest to assume a strong political character."[235]

Given the tight schedule for the 1990 NAEP mathematics test, some analysts cautioned that perhaps NAGB should not try to set achievement levels at this time. But NAGB believed it was important to proceed immediately, although it agreed that the results should be viewed as developmental and provisional. An advisory panel of sixty-three judges was appointed in June 1990 and met in Vermont on August 16–17, 1990, to make three rounds of ratings "indicating what proportion of students at each achievement level ought to answer each particular answer correctly." A technical advisory committee then met and revised the procedures. Thirty-eight of the sixty-three judges met again in Washington on September 29–30, 1990, and participated in two additional rounds of ratings. Another, smaller meeting of eleven judges occurred about six weeks later, in which they wrote descriptions of the three achievement levels and mailed the document to panel members for approval (forty-five approved and eight expressed disagreement with part of or the whole document). The board also received comments from the public as well as its own evaluators and convened all-day meetings in four states to discuss the materials. At the May 1991 meeting, the board voted 19–1 to adopt the proposed achievement levels and accepted the recommended percentage of correct answers for each level.[236]

Throughout the standards-setting process for mathematics in 1990, critics continued to question the value of pursuing this undertaking. Frank Betts, director of the Curriculum and Technology Center for the Association for Supervision and Curriculum Development, protested: "I believe there is a very real potential for harm in reporting three levels of achievement."[237] Herbert Rosenthal, former deputy director of the Central Park East Secondary School in New York City, believed the math standards were based on outmoded ideas and would do little to improve classroom practices.[238]

Some other individuals and organizations also challenged the speed and quality of the standards-setting process. The National Council of Teachers of Mathematics believed the process had been too hasty and was technically flawed. Gregory R. Angrig, president of ETS, told NAGB: "I think you can do a better job in 1994, and an even better job in 1996. I want you to get off to a good start. The danger is, if you move too fast in the wrong way, you'll lose what you are trying to accomplish."[239] Finn, however, replied: "I agree it would be good to take time to do things well. But I am also mindful of the adage, 'the perfect is the enemy of the good.' If we do not get baseline data until 1995, we may be sacrificing something else—the sense of urgency for national improvement."[240]

Disagreements over the setting of the 1990 mathematics performance standards surfaced in a variety of different reports and meetings.[241] But one of the most contentious and bitter controversies grew out of a very small NAGB evaluation contract to assess its own standards-setting process. The board hired Aspen Systems as a logistical service intermediary for contracting an independent evaluation of the achievement-levels setting for the 1990 mathematics evaluation. Aspen Systems subcontracted with three well-known researchers to carry out the evaluation—Daniel L. Stufflebeam and Michael Scriven from Western Michigan University and Richard M. Jaeger from the University of North Carolina at Greensboro. The initial $11,000 contract was for work done before December 1, 1990; another $7,000 was added for additional work later. The researchers were to participate in and assess all phases of the standards-setting process. The investigators submitted interim reports to NAGB on November 7, 1990, and January 14, 1991, that provided suggestions for improving the ongoing process. A third report was delivered on May 5, 1991, and warned against releasing the results without adequately warning the public about the conceptual and technical shortcomings in the standards-setting process.[242]

Stufflebeam and his colleagues became frustrated that their recommendations in the third interim report had not been adequately incorporated by NAGB. When they issued a draft summative evaluation on August 1, 1991, it was highly critical of the entire standards-setting process. The document was marked "confidential" and "do not reproduce or circulate." But the authors distributed copies to approximately thirty-nine individuals (some of whom were policymakers rather than testing experts) for comments without first seeking authorization from NAGB—an action that was quickly denounced by NAGB.[243]

Stufflebeam, Jaeger, and Scriven catalogued and analyzed the problems in NAGB's standards-setting process in the draft report. They concluded that "the technical difficulties are extremely serious and not mere academic complaints about finer points of test design and interpretation. Consequently, the resulting standards, which are due to be released in spite of the project's technical failures, must be used only with extreme caution."[244] They drew two major conclusions from the technical difficulties associated with the project:

> 1. *These standards and the results obtained from using them should under no circumstances be used as a baseline or benchmark* against which future changes in performance are to be measured as representing progress or its absence. To do so would likely cause substantial errors of policy and massive waste of resources. Substantial amounts of real progress would not be credited; substantial amounts of pseudo credit could well be cheered. In the very tight time line for achieving the White House education goals for this century, this kind of mistake could make the difference between success and failure.
>
> 2. *The procedures used in this exercise should under no circumstances be treated as a model* in other subject matter areas. They were a reasonable first ever attempt to use the Angoff procedure to set three achievement levels on an existing test; they would be ridiculous as a repetition, in the light of the problems that turned up and that must be solved before moving ahead. In the body of this report we have specified design issues that must be resolved before NAGB can confidently proceed to set achievement levels for future NAEP assessments. We believe these can be addressed in less than a year. Proceeding to replicate the process reviewed here would essentially eliminate the credibility—and almost certainly much of the utility—of everything that is built on the results of the replication.[245]

Stufflebeam and his associates recommended that NAGB "suspend its level-setting effort and redesign it so that it can produce technically defensive results." They also urged NCES to "delay funding additional levels-setting projects until there is a sound technical basis for additional projects."[246] But the researchers did not just point out the conceptual and technical shortcomings in the standards-setting process; they also questioned the technical competence of NAGB and recommended that the Congress reconstitute the board:

> However, the composition of NAGB is problematic. While it appears to meet the Congressional requirement to be bipartisan and broadly representative of local, state, and national stakeholders, it includes too little expertise from the psychometric and evaluation communities to ensure that it will perform its policy making and test design responsibilities in accordance with the published and demanding standards of the field of educational and psychological measurement....
>
> The problems attending NAGB's lack of technical expertise seem to present a serious policy issue rather than an operational issue. There are compelling reasons why NAGB should be sustained, provided it can effectively increase the interpretability and appropriate use of NAEP results. However, it probably cannot fulfill its responsibility if the technical community is not represented on the Board at a level equivalent to that of the NAEP user groups. During our study of this inaugural levels-setting project, it was apparent that the few technical people on the NAGB staff and even their superb consultant, Dr. Ronald Hambleton, were powerless to do what their technical expertise told them should be done, because they were led by a politically oriented and effective Executive Director and only one technically oriented counterpart on the Board. In retrospect, it seems clear that the inaugural project's chaotic nature, including recycling the project three times, was a direct function of the Board and its Executive Director making technical design decisions that they were not qualified to make.[247]

NAGB was outraged by the draft report and believed it was politically motivated as well as technically and factually inaccurate. Richard Boyd, chair of NAGB, almost immediately sent a copy of the draft report and his own critical response to it to the board members.[248] Simultaneously Boyd, Mark Musick (vice chair), and Michael Glode (committee chair) wrote an official response to those who had received a copy of the draft report:

> It is our considered judgment that the draft report is so thoroughly flawed as to be unsalvageable. Stufflebeam, et al., misperceived the Board's role, which essentially is judgmental, not technical. Further, they misperceived their own role, which was technical, and made it plainly political with this report. Therefore, we will not invest additional resources in attempting to rectify its shortcomings. Instead, the draft report, the response of the National Assessment Governing Board to the draft, and all original documentation for the project will be shared with interested parties.
>
> Editing alone cannot correct the egregious errors and misstatements of fact, or undo the purposeful lack of objectivity permeating the entire document. To purge the hyperbole, eliminate the innuendo, rectify the misperceptions and correct the deficiencies would be far too comprehensive a task to undertake in the time available. In addition, the authors have failed to follow the generally accepted standards in their field....
>
> In the final analysis, the Board made a judgment that it believes is sound and defensible. As a matter of fact, the Board followed many of the recommendations made during the formative stages of the evaluations conducted by the authors of this report. However, much of the tone, tenor and substance of the recommendations contained in the draft summative report are inconsistent with those earlier recommendations, and new objections are raised in this report that were never hinted at previously. The Board accepted many, but not all, of the authors' earlier recommendations. For that matter, the Board did not accept all of the recommendations of any other person, group or organization. It made its own judgments as it was obliged to do.[249]

After consultation with the Office of General Counsel of the Department of Education, NAGB took steps to terminate the subcontract with the three researchers—only to discover later that the contract had already been officially completed and therefore could not be terminated.[250] Included with these letters was NAGB's more detailed 23-page reply, "Response to the Draft Summative Evaluation Report...."[251] Stufflebeam and his colleagues submitted a revised, final report on August 23, 1991, but NAGB did not acknowledge receipt of that document.[252]

The controversy over the draft report by Stufflebeam and his associates escalated. *Education Week* ran an article with the provocative headline: "NAEP Board Fires Researchers Critical of Standards Process." The text of the article was balanced, but it attracted considerable attention to the issue:

> The governing board of the National Assessment of Educational Progress has fired a team of researchers that prepared a critical evaluation of the board's process for setting achievement levels for use in reporting results of NAEP's 1990 mathematics assessment.
>
> In a letter to researchers and policymakers who had received a draft copy of the final report, officials of the National Assessment Governing Board called the draft "so thoroughly flawed as to be unsalvageable." In addition to numerous "egregious errors and misstatements of fact," the officials stated, there is a "purposeful lack of objectivity permeating the entire document." The officials also charged that the reviewers had committed a "political act" by submitting the draft to policymakers, such as Congressional aides and Gov. Roy Romer of Colorado, rather than just to technical experts who could make informed comments.
>
> "It was clear from the beginning they didn't believe in achievement levels, they didn't want levels set," Roy E. Truby, executive director of the N.A.G.B., said in an interview. "It was even more clear they don't like a N.A.G.B.-type board." He said the board was terminating the evaluators' services because it was unwilling to invest additional funds in editing the final report.
>
> Richard M. Jaeger, a professor of education at the University of North Carolina at Greensboro and a member of the evaluation team, responded that the board's charges were "absurd," and he took offense at the charge that the team was biased and had failed to follow the standards of education program evaluation. The chairman of the review panel, Daniel L. Stufflebeam, director of the evaluation center in the college of education at Western Michigan University, wrote the evaluation standards, Mr. Jaeger noted. These charges are "code words for saying they didn't like our conclusions," Mr. Jaeger said. "This is a case of not liking the message and acting to kill the messenger."
>
> Other researchers familiar with the incident voiced outrage at the board's action, and warned it could threaten the credibility of NAEP....Mr. Truby denied that the board was attempting to squelch the evaluators' report, and noted that he was making available the draft report, along with the board's rejoinder, to all interested parties. Mr. Jaeger said that, despite the firing, the panel had submitted its final report to the N.A.G.B., and added that he hoped others would read it and come to their own conclusions. "I'm hopeful people who are less passionate about the argument will evaluate the evidence on its merits," Mr. Jaeger said.[253]

John F. Jennings, counsel to the House Education and Labor Committee, indicated that the committee planned to ask the General Accounting Office (GAO) to reexamine the standards-setting process. Jennings stated that "we want them [GAO] to make a judgment about whether the Stufflebeam team was correct on their points, or whether N.A.G.B. was correct in their rebuttal."[254] On October 7, 1991, Congressman William D. Ford (D-MI), chair of the Committee on Education and Labor, and Congressman Dale E. Kildee (D-MI), chair of the Subcommittee on Elementary, Secondary, and Vocational Education, asked GAO to review how NAGB had established student performance standards.[255]

After an interim response on March 11, 1992, GAO issued its final report on June 23, 1993. The title on its publication clearly indicated its dissatisfaction with the work of NAGB—*Educational Achievement Standards: NAGB's Approach Yields Misleading Interpretations.* The report went on to say:

> GAO found that NAGB's 1990 standard-setting approach was procedurally flawed and that the interpretations that NAGB gave to the resulting NAEP scores were of doubtful validity. While the scores selected represent moderate, strong, and outstanding performance on the test as a whole, GAO concluded that they do not necessarily imply that students have achieved the item mastery or readiness for future life, work, and study specified in NAGB's definitions and descriptions. The difficulties evident in NAGB's 1990 achievement levels resulted in part from procedural problems but also from the effort to set standards of overall performance (how good is good enough) that would also represent standards of mastery (what students at each level should know and be able to do). NAGB improved its standard-setting procedures

> substantially in 1992, but the critical issue of validity of interpretation—an issue in NAGB's approach—remains unresolved. GAO therefore concluded that NAGB's approach is unsuited for NAEP.
>
> GAO identified several alternative approaches that could be used to establish standards for overall performance on a NAEP test. However, any approach that sets standards purporting to measure mastery of particular subject content will be difficult to use with NAEP as it is currently designed.
>
> GAO found that in the case of the achievement levels, NAGB designed and implemented its approach without adequate technical information. In two other cases, however, NAGB made better use of such information. GAO concluded that NAGB's composition, procedures, and relationships with the Department of Education are inadequate to ensure that policy guidance to NAEP will be technically sound.[256]

The GAO report, like the Stufflebeam, Jaeger, and Scriven evaluation, questioned NAGB's technical expertise:

> GAO concluded that NAGB's strength lies in its broad representation, not in its technical expertise. However, the law assigns NAGB responsibility for some functions that are clearly technical and for others that have both technical and policy implications. From examining three decisions, GAO found that when NAGB recognized an issue as clearly technical, it sought and used expert technical advice in policy planning and sometimes in implementation. However, NAGB initially considered the setting of achievement levels a policy function that it itself could perform with minimal technical support and did not appreciate the importance of verifying the validity of its score interpretations. NAGB's governance structure and procedures neither ensure that technical issues will be recognized nor require that technical considerations be addressed early in the policy formation process. GAO thus concluded that there is substantial continuing risk that NAGB may give NAEP technically unsound policy direction.[257]

GAO concluded its evaluation with several specific recommendations:

> Since the current NAGB approach to setting standards has yielded unsupported interpretations of NAEP scores, GAO recommends (1) that NAGB withdraw its instructions to NCES to publish 1992 NAEP results primarily in terms of levels of achievement, (2) that NAGB and NCES review the achievement levels approach, and (3) that they examine alternative approaches.
>
> To strengthen NAGB's capacity to give sound policy direction, GAO recommends that NAGB (1) obtain NCES review of proposed policies; (2) conform to its own policy of prescribing policy ends, not technical details; and (3) nominate for the testing and measurement positions on NAGB persons who are trained in the design and analysis of large-scale educational tests. GAO also recommends that the Congress clarify what it intends NAGB to do with respect to achievement goals and review the division of responsibilities between NAGB and NCES, with a view toward concentrating NAGB's efforts on the representational functions for which it is well designed.[258]

GAO encourages the federal agencies it evaluates to respond to its critiques. NCES generally agreed with much of GAO's analysis, but differed on a few points. NCES did endorse the idea of clarifying the roles of NCES and NAGB:

> We concur that statutory clarification of the important roles that NCES and NAGB have to play in the NAEP project could serve a constructive purpose, and we will also consider this as we develop our reauthorization proposal. NAGB is well suited to provide broad policy advice by representing the many constituents served by the NAEP project. NCES is well suited to provide the operational and technical expertise needed to conduct a complicated survey like NAEP. Both functions are needed in order to ensure that the assessment data are technically valid and reliable and, at the same time, policy relevant and worth the expenditure of considerable public funds.[259]

NAGB, on the other hand, strongly disagreed with the GAO report. The board emphasized four major points:

> National Assessment results should be reported primarily in terms of challenging standards that help the nation determine "how good is good enough." The conventional practice of simply comparing one group of students to another is no longer adequate. GAO makes no compelling argument for returning solely to the older methods of reporting by means, percentiles, and "benchmarks."
>
> The Board and numerous other groups believe that achievement levels can properly be used to report results on the National Assessment. We reject the

> argument that trying to set standards on NAEP is "conceptually flawed." We reject GAO's recommendation that the 1992 achievement levels be withdrawn. The GAO report is unbalanced and misleading. Many of its assertions are undocumented; much of its analysis is flawed. The GAO report is out-of-date. It focuses on the achievement levels for 1990—indeed, mostly on the first phase of the process for setting them which did not form the basis for the levels actually adopted. It gives relatively little attention to the standard-setting process for 1992 and fails to recognize the improvements made.[260]

NAGB pointed out that the 1992 achievement-setting process was much better than the earlier one, especially since it "was conducted under a $1.5-million contract by American College Testing (ACT), which has extensive experience in standard-setting in many fields."[261] Moreover, NAGB rejected GAO's negative views of the organization's technical capabilities and achievements:

> The Governing Board agrees with GAO about the importance of securing technical advice, and has done so regularly in regard to achievement levels, as well as in its other work. However, because of the wide impact of NAEP, the assessment should be guided by an independent, widely-representative policy-making board—not a closed circle of federal officials and technicians.[262]

As with the Stufflebeam, Jaeger, and Scriven draft report, the GAO criticisms of NAGB were widely publicized among educators and policymakers. When the GAO interim report was issued, the headline in *Education Week* read: "G.A.O. Assails Standards-Setting Process for NAEP."[263] When the final GAO evaluation was issued the following year, *Education Week* announced: "G.A.O. Blasts Method for Reporting NAEP Results."[264] Particularly disturbing in the latter story was the reporter's claim that GAO had concluded that the NAGB standards-setting process was "fundamentally flawed."[265] Mark Musick, chair of NAGB, replied in a letter to the editor of *Education Week*:

> Unfortunately, there is a serious mistake in the lead of your July 14, 1993, article on the U.S. General Accounting Office report about achievement standards for the National Assessment of Educational Progress—even though the story itself is generally fair and balanced ("G.A.O. Blasts Method for Reporting NAEP Results").
>
> Contrary to the phrase in quotation marks, the G.A.O. report never says that the standard-setting approach used by the National Assessment Governing Board is "fundamentally flawed." The report certainly does criticize the Governing Board. But the central point of its criticism is actually quite limited. As the title of the G.A.O. report indicates ("N.A.G.B.'s Approach Yields Misleading Interpretations"), it deals primarily with a question of reporting and interpretations—not with the broad issues of whether achievement standards can be set on the National Assessment of Education Progress or the basic process for setting them.[266]

This clarification notwithstanding, concern that the standards-setting process was "fundamentally flawed" reappeared when the National Academy of Education (NAE) issued its in-depth evaluation a few months later.

The Department of Education had been instructed by the 1988 legislation to conduct an independent evaluation of TSAs. NCES commissioned an NAE panel to undertake that task. As criticisms of NAGB's standards-setting process mounted, NCES decided to give the NAE panel the added responsibility of evaluating NAGB's efforts to establish national performance standards. The NAE panel released its report on achievement levels in September 1993.[267]

The NAE panel reviewed the earlier evaluations of the standards-setting process and observed:

> The previous evaluations made several major criticisms. For example, the judgment tasks required by the modified Angoff process were found to be difficult and confusing; the NAEP item pool was not adequate to reliably estimate performance at the advanced levels; the set standards seemed highly dependent on the particular sample of judges; appropriate validity evidence for the cutscores was lacking; and neither the descriptions of student competencies nor the exemplar items were appropriate for describing actual student performance at the designated achievement-level cutscores. All the evaluation studies concurred that the achievement levels, as constructed, were not appropriate for reporting NAEP results.

> The Governing Board was responsive to many of the concerns of its evaluators, and it did designate the 1990 achievement levels as a trial effort. However, NAGB remained committed to delivering final achievement levels for use in reporting 1992 results, and, consequently, advice that suggested the need for significant additional data collection or a fundamental rethinking of the achievement-level-setting process was not followed.[268]

Rather than just reconsidering the methodology and findings from the previous evaluations, the NAE panel saw itself adding a new dimension:

> Importantly, the Panel's report of the 1992 standards-setting effort is based on new evidence entirely independent of results of earlier evaluations. To the extent that the conclusion of past evaluations are similar to those of the present evaluation, the earlier reports lend additional weight to conclusions in this report.[269]

The NAE panel concluded that the achievement levels-setting process was "fundamentally flawed:"

> The process of developing achievement levels involved two distinct tasks: (1) creating subject-specific descriptions for each level, and (2) identifying cutscores. In both reading and mathematics, the "initial" achievement-level descriptions created by the participants in the level-setting meetings were judged to be inadequate by subject-matter specialists and were substantially revised at a later date. The revisions caused a serious validity problem, however, because the achievement-level cutscores were never reset to correspond to the new desciptions.
>
> The reading process evaluation documented one of the reasons for the inadequacy of the initial descriptions. [Achievement-level-setting] panelists were unfamiliar with the NAEP Reading Framework and therefore used personal experience and opinions to develop the descriptions and to make item judgments rather than following the framework.
>
> The process used to set the 1992 cutscores in reading and mathematics was judged to be indefensible because of the large internal inconsistencies in judges' ratings. The [NAE] Panel's analyses showed that judges could not maintain a consistent view of what a student at the borderline of each achievement level should be able to do. In some cases the internal inconsistencies were huge, with judges setting cutscores for the same level that differed by the equivalent of four to eight grade levels simply as a function of considering different item types. The modified Angoff process also did not facilitate the development of consensus. Differences among judges' ratings were large even at the end of a three-round process. Based on its analyses, the Panel concludes that the Angoff procedure is fundamentally flawed for the setting of achievement levels.[270]

The panel also found that "the weight of evidence suggests that the 1992 achievement levels were set unreasonably high." Overall, the group concluded "that flawed achievement levels would not enhance the interpretability of NAEP and might, in fact, jeopardize other national efforts to develop content and performance standards and might harm the credibility of NAEP."[271]

While the NAE panel reaffirmed its belief in the potential value of voluntary national standards, it expressed disappointment in what had been done to date:

> The members of the Panel strongly affirm the potential value of voluntary national standards that exemplify challenging curricular and performance expectations. However, the standards set must be defensible in order to ensure that assessment data and national education policy based on the standards are sound. Given the problems noted above, the Panel does not believe that the process by which the 1990 and 1992 achievement levels were set can be defended. In the Panel's judgment, setting credible performance standards is a long term process—standards cannot be set in 3 days nor in 3 months.[272]

The NAE panel then made eight short-term recommendations for improving the process of setting achievement levels:

1. Discontinue use of the Angoff method.
2. Discontinue reporting by achievement levels as used in 1992.
3. Invite content experts, business leaders, and standards committees to comment on the meaning of NAEP results and desired performance standards.
4. Publish achievement levels in 1994 separately from the official NAEP reports and report these as draft or developmental.

> 5. Use 1990 and 1992 percentile scores to monitor achievements in future assessments.
>
> 6. Use international comparisons to set benchmarks for U.S. performance.
>
> 7. Work with the National Education Goals Panel to develop a way to use NAEP results to measure progress over the decade of the 1990s.
>
> 8. Implement within-grade score reporting.[273]

The panel also made six long-term recommendations for creating national content standards:

> As national content standards are developed and certified, the Panel believes it is imperative that performance standards on NAEP be linked to them. This will be a time-consuming process. The Panel also believes that the development of such performance standards requires a knowledge base for understanding the meaning of various levels of performance. A knowledge base of this sort cannot be developed quickly enough to be available for the next assessment cycle. For these reasons, the Panel believes that the Governing Board must also take a long view as it seeks to establish performance standards. With this perspective in mind, we turn to the Panel's long term recommendations.
>
> 1. Develop content standards and performance standards in an iterative process.
>
> 2. Establish a standing subject-matter panel for each subject area.
>
> 3. Address important conceptual issues.
>
> 4. Empirically evaluate achievement levels before making them operational.
>
> 5. Recognize the need for a multiyear process for the development of performance standards.
>
> 6. Provide for a stable basis for comparison as well as for evolutionary change.[274]

Although the NAE panel was critical of NAGB's achievement levels-setting process and results, it ended on a hopeful note for the future:

> The Panel believes that a defensible procedure for setting performance standards is well within reach, due largely to the pioneering efforts of NAGB, its contractors, and the many evaluators of the 1990 and 1992 NAEP assessments. The Panel looks forward to the promulgation of rigorous and defensible achievement levels for NAEP, but cautions that it may take some time to establish them. To assist in reaching that objective, the Panel has recommended criteria and procedures for improving the interpretability and usefulness of NAEP reporting, for grounding NAEP in emerging national content standards, and for assuring continued credibility of NAEP as an essential indicator of achievement in American education.[275]

NAGB responded by commissioning several papers to comment on the NAE panel's criticisms and recommendations. Michael Kane, professor of kinesiology at the University of Wisconsin, wrote:

> I think that the evidence provided in the NAE report and in the studies commissioned by the NAE Panel do not provide adequate support for the strong conclusions in the report. The conclusion that the Angoff procedure presents judges with an unmanageable task is based on unwarranted and unreasonable assumptions about what the Angoff procedure is designed to do. The NAE studies attack a straw man when they claim that the Angoff procedure is, "fundamentally flawed," because the ALS [achievement level-setting] panelists exhibited variability in their ratings for different items and because the panelists did not achieve consensus over the three rounds of the rating process.
>
> The conclusion that the standards that resulted from the ALS process are unreasonably high is based mainly on the results of the contrasting-groups study, which has, I think, serious problems. Taken as a whole, the collection of studies examining the reasonableness of the cutpoints give conflicting results. The international comparisons suggest that the cutpoints may be too low. The comparison with the Kentucky system suggest that the cutpoints are about right. The comparison with AP [advanced placement] results suggest that the advanced level at 12th grade may be too high. The comparison with SAT is ambiguous, because we do not have any clear criteria for what should be considered an advanced performance on the SAT. As a group, these studies are a mixed bag and certainly do not support a strong conclusion that the standards are too high.[276]

Gregory J. Cizek, then an assistant professor of educational research and measurement at the University of Toledo, basically agreed with Kane's criticisms of the NAE panel's report and raised some additional issues:

> The NAE Report on the NAEP achievement levels-setting process utilized by NAGB in 1992 contains some positive evaluations and recommendations, as well as negative judgments about the process utilized by NAGB in setting performance levels. However, in my opinion, the report provides an overwhelmingly and overly negative description of the NAEP levels-setting process—a view that is not supported by evidence available for the NAE Report.
>
> In summarizing the results of my review, it is my opinion that the conclusions of the NAE Report: 1) rely on the input of researchers who do not possess relevant expertise in the area of standard setting; 2) do not derive from the application of accepted evaluation guidelines, criteria, or procedures; 3) are presented in a systematically unbalanced manner; 4) are based upon research studies that were not particularly well-suited to answering the questions of interest; and 5) lead to recommendations that would substantially harm the credibility and validity of the National Assessment of Educational Progress.
>
> However, despite the identification of these serious flaws, it should not be concluded from the above evaluation that the NAE Report is without merit. The NAE Report identified issues associated with the levels setting process that warrant further investigation, and issues related to NAEP item development and scaling that are problematic. It can be said that the levels-setting process is not without residual difficulties and drawbacks. On the contrary, because the nature of all standard setting is judgmental, all standard-establishing procedures can be refined and improved. It is unlikely that any process could be designed and implemented in such a way as to be beyond reproach.[277]

NAGB discussed at considerable length the NAE panel report and the criticisms of it by Kane and Cizek.[278] As a result, board members had an appreciation of the contested nature of the discussions over the standards-setting process. But those outside the evaluation field often received a more one-sided impression of the ongoing debate—especially because the report from the prestigious and highly respected NAE was published and widely distributed (and frequently mentioned in the education media) while the short working papers by Kane and Cizek received very little attention and were not circulated broadly. Indeed, while many education policymakers had at least heard of the NAE panel's general criticisms, most were probably not even aware of the comments of Cizek and Kane.[279]

NAGB and NCES sponsored a Joint Conference on Standard Setting for Large-Scale Assessments in October 1994 to examine the technical and policy issues related to setting student standards. The meeting was productive and useful, but no consensus could be reached among the expert participants on many of the key issues:

> It became clear at the conference that standard setters continue to disagree about many aspects of their work. No method of setting standards is universally accepted. The Angoff method, which has been the most widely used means of setting standards, was characterized as "fundamentally flawed" by some authors and defended by others. Not all authors agreed that the use of standards would be beneficial, even if the standards had been appropriately set. Authors elaborated on the difficulties of setting standards, noted the legal vulnerabilities of standards, discussed problems in interpreting the results of using standards, and failed to reach consensus on a number of controversial issues.[280]

James Popham, moderator of the final session of the three-day conference, summarized some of the main themes. Although no agreement could be reached on the appropriateness of using the Angoff method, he felt that most of the experts seemed to support its use—although there was also strong disagreement from others:

> Methods for setting standards may be centered either on overall performance of examinees or on judgments about particular test items. Also, different approaches may be needed for multiple-choice tests and for essay or performance items that can receive a range of scores (the so-called polychotomous items). Most experts in standard-setting believe the widely-used Angoff method of aggregating item judgments is not fundamentally flawed and that the panels it convenes can make the judgments involved.[281]

Popham went on to remind the conference participants that we must be careful not to expect too much of any standard-setting process and appreciate some of the progress that has been made:

> The road to perfection in standard-setting (or anything else) is paved with self-flagellation. We should not be too hard on ourselves or look for a level of precision and accuracy that is not attainable by normals. In fact, standard-setting on examinations is far sounder and more sophisticated now than it was a decade ago. If we proceed in a reasonable, professional, and rational way, we can come up with standards that will be accepted. These standards can be defended against critics and lawsuits, and no judge will rule against them.[282]

Throughout these debates, NCES found itself in a complex and sometimes difficult situation. On the one hand, NCES accepted in principle the value of setting achievement levels. Thus, in his reply to the GAO report on March 25, 1992, Emerson Elliott, at that time the acting assistant secretary of OERI, commented:

> While the General Accounting Office (GAO) report deals primarily with technical aspects of NAGB's actions to set achievement levels, in fact, the concept of performance standards involves much more than that. Any attempt to establish performance standards raises questions of substance: what it is that we want American students to know and be able to do, and how well we expect them to do it. Performance standards also raise questions of public policy: whether our national assessment should lead, or should follow, student learning progress, and how we decide, as a nation, what the standards should be. The Governing Board is attempting to set performance levels to challenge American students. The National Education Goals and the Administration's legislative proposal, Goals 2000: Educate America Act, support that position. On the other hand, the Governing Board also supports gathering high quality data on trends in student performance. The task of balancing these two purposes is challenging, but necessary. The GAO report, however, appears only to support the limited trend monitoring role for the National Assessment. We believe that each of these roles, properly executed, can serve a constructive purpose in informing the public.[283]

On the other hand, the agency was uncomfortable with some of the specific ways in which the performance standards had been set—especially as testing experts and well-respected educational organizations such as NAE criticized the standards-setting process and questioned the validity of the results. Many professionals at NCES would have preferred that NAGB's achievement levels be issued separately from the regular NCES publications, as had been the case with the reporting of the 1990 math assessment.[284] NAGB, however, wanted the achievement levels to be a regular part of the NCES reports on NAEP, as this would give those standards increased visibility and added legitimacy in the eyes of many educators and policymakers.[285] The compromise was to publish the 1992 reading achievement levels in the regular NCES publication, but only after the presentation and analysis of the NAEP composite reading proficiency scale (set to range from 0 to 500). NCES alerted readers to the still-developmental nature of the reading achievement levels and cautioned that new standards-setting procedures might be used in the future:

> The 1992 NAEP *Reading Report Card* marks a continuation of the attempt by NAGB and NCES to shift to standards-based reporting for NAEP. For reading, a transition is being made with the 1992 assessment to report NAEP results by achievement levels that describe how much students *should* know. The impetus for this shift lies in the belief that NAEP data will take on more meaning for the public if they show what proportion of our youth are able to meet standards of performance necessary for a changing world.
>
> Because the progress of setting NAEP achievement levels centers on the descriptions of what students *should* be able to do, it is important also to examine whether students actually meet those expectations for performance. For the 1992 reading assessment, a modified anchoring process was used to examine actual student performance at the achievement levels and describe what they *can* do as demonstrated by their assessment responses. NCES realizes that modifications and improvements may be necessary in the future as current achievement-level procedures are evaluated and new approaches to standards-based reporting are developed by the various parties involved in systemic education reform.[286]

Interestingly, all four of the evaluation studies cited in a footnote had all been critical of the achievement-setting efforts.

NCES followed generally the same reporting procedures for the 1994 reading assessment as it had two years earlier.[287] This time, however, both supporters and critics of NAGB's standards-setting procedures were mentioned in the footnotes. The 1994 report also stated that "the Commissioner of NCES has judged that the achievement levels are in a developmental status."[288]

As NAEP and NAGB were being considered for reauthorization in the early 1990s, some key House Democrats were upset with the ways in which NAGB had set achievement standards and dealt with its critics. The House had accepted the value of NAEP and in May 1993 had unanimously agreed to extend state-level math and reading assessments the following year for the fourth, eighth, and twelfth grades.[289] But the House Subcommittee on Elementary, Secondary, and Vocational Education voted to eliminate NAGB. *Education Week* concluded that "congressional criticism of the [NAGB] board stems primarily from its efforts to set achievement levels to characterize student performance."[290] The article detailed the three recent criticisms of NAGB's standards-setting process and quoted Jefferson S. McFarland, a House staff member, who said: "The subcommittee does not feel a board like the NAGB is necessary and, frankly, we've not been real pleased with its performance.... The NAGB's response to the criticism has been to provide their own critiques of the criticisms, and it just isn't constructive."[291]

Defenders of NAGB pointed to the continued need for an independent group to oversee NAEP. Mark Musick, chair of NAGB, criticized the subcommittee's proposed abolition of NAGB:

> By disbanding the National Assessment Governing Board, the subcommittee plan would exclude the interests of states and other education stakeholders in the governance of NAEP. It would end broad-based public accountability and instead concentrate power in a single federal official [the Commissioner of Education Statistics]. It would threaten the development of high student-performance standards on NAEP, which require the mandate and commitment that only a widely representative board can give.[292]

The National Governors' Association joined in support for the continuation of the board. It passed a resolution that stated that NAGB "is broadly representative of state and local interests, insures public accountability, and maintains the appropriate federal/state partnership in education decision making including the establishment of national, not federal, performance goals for reporting assessment results."[293]

The setting of student achievement standards received a unanimous endorsement from the Education Information Advisory Committee (EIAC) of the Council of Chief State School Officers in May 1994. The membership of EIAC included state assessment directors who relied heavily upon NAEP and were quite familiar with its strengths and weaknesses. EIAC observed that:

> Standard-setting is not a science. There is no one agreed-upon method of approaching the task. Most of the work done in setting achievement levels had been done in establishing cut-scores for competency tests (e.g., high school graduation tests). While such approaches can be used with NAEP and, indeed NAGB did this, there is no requirement that this approach be used.... States were invited to participate in the Trial State Assessment on a voluntary basis. One attraction of TSA was the possibility of obtaining information of the degree to which students were meeting achievement levels. Goals 2000 calls for reports describing the same thing as do many of the individual state programs. The nation certainly is committed to this direction.
>
> NAGB faces an impossible situation if the accusation is made that the standards being set are not accurate or appropriate when there are no guidelines as to what would be an acceptable approach. NAGB has taken great pains to improve the standard-setting processes, but even the current efforts may be faulted by those who do not want any national standards.[294]

EIAC then passed several recommendations that supported NAGB's standards-setting activities:

> NAGB should continue its efforts to establish achievement criteria using the procedures which it finds are as

> credible as possible, yet feasible.... We do not support the idea that the Reading achievement levels should be reported in an "R and D" fashion. We do believe that the information should clearly be reported as being a policy decision of NAGB. If the Reading achievement data is affirmatively adjudicated, the issue is moot. Otherwise, we recommend that NAGB publish a separate report of the data. We affirm our previous position that achievement levels should be the primary reporting vehicle for NAEP data.[295]

Despite the strong support for NAGB by the nation's governors, many in the Clinton Administration and the Council of Chief State School Officers, the House Education and Labor Committee, on a split vote, reauthorized NAEP but abolished NAGB. When the legislation, part of the 1994 reauthorization of the Elementary and Secondary Education Act of 1965 (H.R. 6), reached the House floor, NAGB was reinstated, but in a much weakened form—the board was reauthorized for only two years, limited to $2 million per year, and could no longer determine what subjects would be covered by NAEP or set the achievement levels.[296]

The Senate, a strong supporter of NAGB in 1988 and 1994, insisted not only on preserving the agency, but even more explicitly on empowering the agency to set student achievement levels. The final law (P.L. 103-227) basically endorsed the Senate's position—although it tried to appease the House Democrats by calling the performance standards "developmental" until the Commissioner of Education Statistics specified otherwise:

> (1) PERFORMANCE LEVELS.—The National Assessment Governing Board, established under section 412, shall develop appropriate student performance levels for each age and grade in each subject area to be tested under the National Assessment.
>
> (2) DEVELOPMENT OF LEVELS.—(A) Such levels shall be—
>
> (i) devised through a national consensus approach, providing for active participation of teachers, curriculum specialists, local school administrators, parents, and concerned members of the general public;
>
> (ii) used on a developmental basis until the Commissioner determines, as the result of an evaluation under subsection (f), that such levels are reasonable, valid, and informative to the public; and
>
> (iii) updated as appropriate.[297]

Unlike the growing consensus on the usefulness of reporting state-level NAEP data, there is still considerable controversy about the wisdom of setting student achievement standards and how NAGB has handled that assignment. The tone of that debate seems to be becoming less contentious and divisive than in the early 1990s. A recent Congressional Research Service report, for example, summarized the current situation as follows:

> The selection of performance levels by NAGB, as applied to NAEP test results beginning in 1990, was initially somewhat controversial, leading to substantial debate over whether the benchmarks were reasonable or appropriate. The benchmarks have tended to be somewhat challenging, with the result that relatively few pupils have been determined to meet "proficient" or "advanced" levels of achievement on various NAEP tests to which the performance standards were applied. However, over time, with some revision of the processes by which the performance levels are developed, and input from a variety of sources, acceptance of the benchmarks appears to have increased, or at least debate over their reasonableness has become much less frequent or audible.[298]

In its final summary volume on NAEP in 1997, the NAE acknowledged the steadily growing popularity of the achievement standards, but continued to see the current performance standards as "flawed:"

> Given the growing importance and popularity of performance standards in reporting assessment results, it is important that the NAEP standards be set in defensible ways. Because we have concerns that the current NAEP performance standards (formerly called "achievement levels") are flawed, we recommend that the Governing Board and NCES undertake a thorough examination of these standards, taking into consideration the relationship between the purposes for which standards are being set, and the conceptualization and implementation of the assessment itself. In addition,

> any new standards need to be shown to be reliable and valid for the purposes for which they are being set.[299]

NAGB continued to support the setting of performance standards and worked hard to improve the process. Unexpected difficulties were encountered with the 1996 science assessment and NAGB decided to delay the release of the achievement levels.[300] NCES proceeded with its scheduled report, but the initial release of the science results did not contain the achievement levels.[301] Five months later NAGB released a separate science report that contained the achievement levels.[302]

Evaluations of NAGB and the standards-setting process continue. The 1994 legislation instructed the Secretary of Education to "provide for continuing review of the National Assessment, State assessments, and student performance levels, by one or more nationally recognized evaluation organizations, such as the National Academy of Education and the National Academy of Sciences."[303] The Department of Education selected the National Academy of Sciences (NAS) to conduct the $2-million, three-year study. Faced with a budget shortage, NAGB questioned this expenditure of money. A NAGB subcommittee report stated that the NAS evaluation "is likely to study many of the wrong things, asking many of the wrong questions, and...will result in a report in 1998, too late to be useful."[304] The NAS report is expected to be completed in September 1998 and will provide yet another evaluation of the achievement-setting process as well as other aspects of NAEP and NAGB.

IX Personal Observations about the Future Directions of NAEP and NAGB

Having analyzed some of the past developments in NAEP and NAGB, perhaps this is an opportune time to explore what lies ahead. In previous sections, I have tried to portray as objectively as possible what actually occurred. Now I will offer personal observations about possible future directions for NAEP and NAGB. My views draw heavily from the historical experiences detailed in this investigation; yet the past does not provide us with any obvious prescriptions for the future. Instead, it encourages us to take a broader perspective and to appreciate the various policy alternatives before us today.[305] The number of topics covered in this analysis were limited by space and time considerations. The actual research undertaken in this project, however, was much broader and included inquiries into several areas not specifically addressed in the text. These concluding observations provide an opportunity to comment on some of these other matters by addressing eight major questions about NAEP and NAGB:

1. Should there be a national NAEP?
2. Should there be a state-level NAEP?
3. Do we need student performance standards?
4. Who should oversee NAEP?
5. How effective has the operation of NAGB been?
6. What should be the role of NAGB's professional staff?
7. Do we need a voluntary national test and should NAGB be responsible for developing and overseeing it?
8. What role should NAEP and NAGB play in improving American education?

1. Should There Be a National NAEP?

Before considering the future role of NAGB, a more fundamental question first needs to be answered: "Should there be a national NAEP?" My response is an unequivocal and straightforward "yes." The nation needs an objective, reliable assessment of how well our K–12 students are doing as well as information on long-term trends. Therefore, it is important that we continue to maintain comparable assessments over time even as we work toward aligning the current tests more closely with the newly emerging national and state content standards.[306] Maintaining and improving the national NAEP should be one of our highest priorities; NAGB, NCES, and most other education organizations strongly agree.[307]

The Alexander-James report and the NAE commentators emphasized the importance of having the NAEP assessments cover the entire academic curriculum to address the educational needs of the nation.[308] During the past decade (1989–1998), national NAEP main assessments were administered in eight different areas: arts (eighth grade only), civics, geography, mathematics, reading, science, U.S. history, and writing. But there has been more emphasis on mathematics and science (five tests altogether) and on reading and writing (seven tests altogether) than on civics, geography, and U.S. history (three tests altogether). Although long-term trend data were gathered five times for mathematics, reading, science, and writing, none were collected for civics, geography, or U.S. history.[309]

The relative lack of attention to the social sciences on the national NAEP exams continues in NAGB's proposals for next decade (1999–2008)—although there are some important changes. National NAEP main assessments will be done again on mathematics and science (six tests altogether—one more than during the previous decade); reading and writing (four tests altogether—three fewer than during the previous decade); civics, geography, and U.S. history (three tests altogether—the same as during the previous decade). Long-term trend data for mathematics, reading, science, and writing will be gathered only three times (two times fewer than during the previous decade). On the other hand, twelfth-grade information in world history and economics will be gathered only once, in 2005.[310]

While some efforts are being made to assess the social sciences in national NAEP, less attention still is being devoted to these subjects than to mathematics and science or reading and writing. The decision not to collect any long-term social science trend data is particularly puzzling and disappointing. Given the projected growth of cultural and ethnic diversity in our population during the next decade and the benefits of an educated and civic-minded electorate, perhaps NAGB should reconsider its plans regarding the social sciences.[311] Congress and the public have often expressed their interest in the importance of civics, geography, and U.S. history in the curriculum, but they have also been concerned at times about the controversies surrounding the national history standards developed by the National Council for History Standards.[312] NAGB, which has developed content frameworks in civics, geography, and U.S. history without arousing much controversy, may be able to provide additional leadership in the future, but it should demonstrate even more interest and commitment to the development and assessment of the social sciences than we have seen to date.[313]

Providing additional coverage in the assessment of the social sciences would require additional funding and I am hopeful that the federal government will be able to provide that financial assistance. If those monies are not forthcoming, however, NAGB might consider recommending shifting to the social sciences, if possible, some of its proposed expenditures on the state-level assessments or on the development of the voluntary national tests. Maintaining and improving the national assessment system is so important that care must be taken that other worthy endeavors do not hinder adequate funding for the national NAEP.

2. Should There Be a State-Level NAEP?

Although almost everyone can agree on the need for a national NAEP, there has been much more controversy since the mid-1960s about the wisdom of having a state-level NAEP. Much of that concern was fueled by fear that any systematic, state-level data on student outcomes might lead to unfair comparisons and conclusions about the quality of public schools in different states.

Since the mid-1980s, however, many state governors have viewed state-level NAEP as an important component of their school improvement efforts. Congress created the trial state assessments (TSAs) in 1988 and NAGB worked hard to develop and implement them. One frequently forgotten, but very important success story of the past decade was the widespread acceptance of the value and reliability of state-level NAEP assessments.

BEST COPY AVAILABLE

Given the continued interest in state-level NAEP by the public and policymakers, we should continue to provide those data on a regular basis. NAGB's proposed schedule of staggered fourth-grade and eighth-grade assessments in mathematics, reading, science, and writing at the state level every four years is reasonable in view of the limited funding available.[314] Having the states make in-kind contributions to help defray at least some of the costs of those examinations is not unreasonable, although the federal government might want to ensure greater participation by funding at least some of the state-level assessments entirely.[315]

Some states also might be interested in using NAEP-related tests in such other subjects as civics, geography, and U.S. history (especially in the twelfth grade). NAGB should try to help provide states with the necessary information to either use such assessments as a whole or incorporate portions of them into state examinations as supplements. It is important to keep the disciplines of subjects such as geography and history separate, as stated in the National Education Goal Three, rather than merged into a more diffuse and less useful general category called "social studies."[316] It would be helpful for the federal government to provide additional assistance for an expansion of state-level NAEP, but such assistance should not come at the expense of providing adequate support for coverage of the core academic curriculum in the national NAEP.

Perhaps one way of providing states with more potential resources for conducting NAEP state-level assessments—either more frequently or in more subject areas—is to make available directly to the states some, but not all, of the now-targeted federal education monies for technical assistance. For example, some of the technical assistance monies now allocated directly to comprehensive service centers or regional educational laboratories might be allocated directly to the states.[317] Then states could continue to purchase some of their needed technical assistance from the comprehensive service centers and regional educational laboratories; they might find it more convenient and efficient to use other educational service providers as well.[318] States might also consider using some federal monies to fund any additional state-level NAEP assessments. While one might insist on providing some basic state assessments or technical assistance services through existing federally funded institutions, we might also grant states additional federal funds that could be spent more flexibly to address their own particular educational needs and priorities.

3. Do We Need Student Performance Standards?

The past decade has witnessed intense debates over the setting of NAEP student performance assessments. Some policymakers have questioned the value of developing any student achievement standards, and many testing experts have challenged the particular ways in which NAGB has gone about this task. Although the heated nature of this debate has diminished over time, strong differences of opinion remain among some of the major participants.

Public citizens and many policymakers have expressed strong interest in and support for student performance standards. They believe that the nation needs to know what is expected of our students and how well they are meeting those standards.[319] I agree with the need for setting rigorous, high-level, student achievement standards—both to stimulate educational reforms and to assess how well we are achieving our stated objectives. Simply reporting student achievement scores using an arbitrary mathematical scale that is difficult for the average citizen or policymaker to interpret is not enough.

Setting student performance standards, however, is a much more difficult task than simply agreeing to the need for them. At the outset, we should acknowledge that any such standards involve a high degree of judgment and inevitably lead to legitimate differences of opinion among those involved. Any group setting those standards, then, must have the legitimacy and ability to establish a credible process that considers both public opinion and the long-term educational needs of the nation.

Setting student performance standards involves judgment, but it also is dependent on the strengths and weaknesses of the conceptual and statistical procedures used to provide the necessary technical information for the judges. As we have seen, there is no single, agreed-on way to develop student achievement levels; different procedures often yield quite different results. Those who are to make the judgments on the expected levels of student achievement must be provided with technically sound information based on whatever particular approach is adopted. Moreover, the group that oversees the entire judgment-making process and selects the technical approach to be followed should understand the advantages and disadvantages of the alternative procedures to know what judgments to make about the final results.

Considerable progress has been made in the past ten years in our understanding of standards-setting processes, as well as in our ability to implement them in the national and state-level NAEP assessments. But much still remains to be done. We need additional studies and more expert discussions about the standards-setting process. We need to find better ways of communicating the existence of that complexity to policymakers and to the public so they will better understand and appreciate what NAEP student achievement standards mean. We also need to devote much more time and energy to ascertaining the meaning of the various levels of student achievement for subsequent individual development. For.example, are students who are designated as "proficient" in a subject in the twelfth grade much more capable of doing college-level work in that area than those who were categorized as having only a "basic" knowledge? Is the attainment of a "basic" level of achievement in twelfth-grade U.S. history or civics an adequate background for becoming thoughtful and involved citizens as adults?[320]

4. Who Should Oversee NAEP?

The nation must have confidence that its report card is being filled out in an objective, nonpartisan manner. Since one of NAEP's primary functions is the development of student performance standards, it needs the involvement of a highly respected group that is credible to both policymakers and the public.

An existing, traditional federal agency might provide assistance, but it should not be given responsibility for overseeing NAEP—especially if that organization is not comfortable setting student performance standards, which inevitably involves making value judgments. There have been efforts by particular administrations to influence and control the management and interpretation of NAEP. Therefore, assigning the direction of NAEP to a relatively independent organization like NAGB, which has broad, bipartisan membership capable of making judgments about student performance standards, is a good idea and should be continued.

While policy decisions on the design and development of NAEP should remain with NAGB, NCES should continue to provide technical assistance to and oversight of the NAEP contractors. NCES is well suited to provide technical assistance for the design, fielding, and data analysis of NAEP. The involvement of both NAGB and NCES in working with NAEP does introduce some additional complications and tensions, but it also brings invaluable synergies to the project. Although the overlap of the two agencies creates some inefficiencies, it acts as a valuable informal check on each agency to help ensure that NAEP remains as objective and nonpolitical as possible. In recent years NAGB and NCES have demonstrated increased cooperation and there is every reason to believe that this will continue, at least in the near future.

As Congress deliberates the reauthorization of NAEP and NAGB in the future, it should also keep in mind the presence of other groups that have been interested in NAEP. The Advisory Council on Education Statistics (ACES) was expanded in 1994 at the behest of some members of the U.S. House of Representatives in anticipation that ACES might eventually assume some of NAGB's current responsibilities. Because NAGB has not only survived, but continues to flourish, what should be the role of ACES in the future? Since the National Education Goals Panel (NEGP) is also awaiting reauthorization, what should be the relationship between NAGB and NEGP? NAGB should continue to provide the sole policy oversight for the development and implementation of NAEP. But groups like ACES and NEGP, which are knowledgeable and interested users of these data, should continue to have ample opportunities to make any helpful suggestions about how to improve the operation of NAEP today or how to introduce ways to enhance future assessments. Thus, at the same time that the future responsibilities of NAGB and NCES for the development and oversight of NAEP are being discussed, we should consider the other congressionally mandated advisory panels that have relied heavily on NAEP data for their own work.

5. How Effective Has the Operation of NAGB Been?

Overall, NAGB has effectively overseen the development and functioning of NAEP, although there are some areas in which the board was less effective than others. NAGB members have been active and thoughtful in the discharge of their duties, and they have maintained a balanced, bipartisan approach to most issues before them. The board has acted as an unusually cohesive and openly deliberative body and generally has worked well with federal agencies such as NCES or outside organizations such as the Council of Chief State School Officers. The board expanded state-level NAEP assessments and worked hard to try to persuade Americans of the need for student performance standards. And NAGB appears to have done an admirable job on several important issues that have not been addressed in this report, such as the development of challenging content frameworks.

Only on the difficult issue of setting student performance standards did NAGB encounter much concerted opposition. The board was particularly committed to the development of such standards, reflecting in part the strong desire among the public and policymakers for that information. But critics believed that NAGB sometimes moved too quickly and without sufficient technical preparation or an adequate understanding of the complexity of the standards-setting process. On the other hand, some NAGB members believed that many critics were biased against any student performance standards and focused too narrowly on the technical aspects of the process, thereby minimizing the judgmental nature of the endeavor. Whatever ultimate judgment one might render about this entire process, some important gains have been made during the past decade in the development and acceptance of NAEP student performance standards—but sometimes at considerable cost to the reputation of and support for NAGB among most of the testing community and some policymakers. And although there is ample praise and blame for all sides during this rather heated controversy, one wonders whether the board might not have found ways to more tactfully and effectively address its critics.

The procedures by which board vacancies are filled is also an important and sometimes contested issue. Initially the board was to nominate three individuals for each vacant position and the Secretary of Education would select one or ask for additional nominees from NAGB. This process might have permitted the board to be self-perpetuating to some degree, so the procedure was changed in 1994 to allow the Secretary to select from nominations provided from the outside. At the request of Secretary Riley (a request that might be changed by his successors), the process is still overseen by NAGB. The new system seems to have worked quite well, although some concerns have been raised about what might happen if a new Secretary of Education is not sympathetic to the way NAEP and NAGB have developed during the past decade.

Having the nominations for the board be generated by others in addition to NAGB seems sensible in a democracy, where the reality—or even the appearance—of any self-perpetuating body in charge of something as important and sensitive as NAEP is to be avoided. One might consider developing provisions under which NAGB might be asked either to comment on the nominees or perhaps even veto some of them to help preserve the integrity and independence of the board. At the same time, we should remember that four different Secretaries of Education have been involved in the selection process and the results to date have been reassuring.[321] A strong political culture and tradition has already developed that emphasizes the value of an independent, bipartisan NAGB and encourages future Secretaries of Education to select distinguished and open-minded board members. The culture and atmosphere of NAGB itself also helps new members to leave behind their own, more narrow interests and to strive to maintain the broader goals and values embodied in NAEP.

The length of appointment for a board member was reduced by 1994 legislation from four years to three years and members were limited to serving only two terms. Given the complexities facing new board members and their demanding workload, it seems more reasonable and realistic to go back to four-year appointments with the possibility of the appointments being renewed once. As long as board members discharge their duties in a responsible manner, the Secretary of Education should be encouraged to reappoint them to second terms. Many board members could serve with great distinction for more than two terms, but the idea of limiting the total number of terms served is a good one and should be maintained.

There has been considerable discussion and some disagreement about the need for board members who are more technically oriented. One plausible argument is that such individuals are unnecessary on the board itself because that same expertise can be recruited whenever needed by convening special panels or hiring expert consultants. Others, however, point to the complex technical issues faced by the board and have argued that the presence of more technically sophisticated members is not only useful, but essential. Reviewing some of the more difficult technical issues before NAGB and considering the role that the congressionally mandated technical experts on the board have played, it seems that the slots allocated to them have been well used and should be maintained in the future.

6. What Should Be the Role of NAGB's Professional Staff?

NAGB's small professional staff have played an important role in providing guidance and assistance to the board. Ably led by Roy Truby, the executive director, the professional staff have worked harmoniously and efficiently with the board members and are to a significant degree responsible for much of the effectiveness of the operation. There are some issues in regard to that staff, however, that might be explored in the future.

There has been very little turnover in the professional staff, which has contributed to the stability and effectiveness of the board's work. At the same time, however, the conceptual and statistical work in particular areas such as testing and the social sciences has made important advances. NAGB staff often hear about those changes in the course of their work—but are they receiving the additional training they might need to keep up with those developments in their fields? Does NAGB routinely provide financial assistance and release time for its professional staff to keep up with recent developments? In addition, are the professional staff given the opportunity and encouraged to develop their own contributions and publications?

It appears that NAGB staff are treated in some ways like the professional members of congressional staffs, who are not expected to publish under their own names or write articles or books as part of their regular duties.[322] One consequence of this approach is that Congress has difficulty recruiting and keeping highly trained and ambitious professional staff members.[323] Does this general approach to NAGB's professionals also hinder or limit their careers as well as diminish their standing and interactions with the other scholars and academics they see in their work?

The professional staff effectively provide for the needs of the board, but are they also making a significant contribution to the larger assessment and educational community? For example, NAGB spends a significant amount of money on consultants and contracts to develop and explore ways of improving NAEP; does the office routinely publish those contributions? The failure to publish and distribute more widely the useful critiques by Cizek and Kane of NAE's analysis of NAGB's student performance standards hurt the agency by making it appear that the board was ignoring NAE's criticisms and recommendations without a proper basis. Unless people attend NAGB's quarterly meetings, how do they know about the advances in the field sponsored by NAGB?

One might contrast NAGB's limited publications program with that of NCES. NCES itself has in recent years increased its rate of publication to include more of its important scholarly and policy contributions aimed at quite diverse audiences. It will cost more money to publish additional materials, but isn't it rather inefficient and wasteful to commission various important studies, which then are not shared more broadly—even through a relatively inexpensive series of working papers? What happens to the studies and analyses performed by the professional staff? Are they presented at various professional conferences and published in the appropriate scholarly and policy journals? Perhaps NAGB should examine more closely how much it spends on the generation of knowledge related to NAEP and what happens to that information after the agency receives it.

The larger question one might raise is whether NAGB and its professional staff are just the developers and overseers of NAEP, or whether they are also expected to be among the intellectual leaders in the field of assessment and educational reform. Since the board develops and oversees NAEP as part of the broader effort to improve American education, one might argue that NAGB should consider itself a major intellectual and policy contributor in educational reform, not just a thoughtful and effective planning unit that develops and monitors NAEP. But if that is to be the case, NAGB will need to hire additional, highly distinguished, innovative experts who might also make cutting-edge contributions to the fields of assessment and educational improvement. The agency and its current professional staff then will need to make more of an effort to share their knowledge and expertise with the broader educational and policy community. And if NAGB chooses not to pursue this more ambitious role or cannot do so in practice, what other groups or organizations should be encouraged to assume those responsibilities and opportunities?

7. Do We Need a Voluntary National Test and Should NAGB Be Responsible for Developing and Overseeing It?

Educators and policymakers in recent years have become preoccupied with debates over the proposed voluntary national tests. Congress has designated NAGB to oversee preliminary development of the tests and subsequent implementation if the full-scale fielding of those tests is approved.[324] Although this analysis did not discuss the issues surrounding those tests in any detail, based on some of my research into the area as well as my own values, I can appreciate the potential need for them. I am concerned, however, about the limited number of subjects to be tested and the long-term, ongoing costs of the tests (although in the short run it makes sense to focus on reading and math). Perhaps we also need to focus more attention on the reliability of the proposed individual-level tests for ascertaining and reporting accurately the extent and nature of an individual's subject-matter knowledge, given the limited testing time available (especially since performance on such an exam may come to have important consequences).[325]

If voluntary national tests are to be administered, NAGB is certainly a reasonable choice for overseeing their development and implementation. NAGB provides the independence and bipartisan orientation that the public and policymakers would like to have for such a sensitive and important assignment. But why wasn't NCES asked to be a partner that might provide invaluable technical assistance and additional assurance to others that this task will be technically sound? Does NAGB alone have the technical knowledge, expertise, and credibility necessary to undertake this important initiative? Given the inevitably contentious issues that are likely to arise, has the board learned enough from its experience setting student performance standards to deal more diplomatically and effectively with its critics? Perhaps Congress might want to revisit the situation and consider whether the effective, but sometimes strained, partnership between NAGB and NCES in the development and implementation of NAEP should continue in some form for the proposed voluntary national tests.

From the perspective of having voluntary national tests, it makes sense to have NAGB develop and implement them. But will NAGB's involvement in those tests greatly diminish its ability to oversee and improve NAEP? The workload of current board members already is very high, and adding such a major additional project may encourage even more individuals to refuse to accept a nomination to that group in the future. The addition of yet another major initiative will stretch the limited resources of the board even further, especially since the amounts of funds and staff provided for the voluntary national tests so far have been minimal. Will the increased focus on the voluntary national tests distract NAGB from making necessary conceptual and methodological improvements to NAEP? And will the inevitable controversies surrounding the voluntary national tests diminish credibility and trust in NAGB and indirectly weaken the development and maintenance of NAEP in the long run? Those of us who believe in the value and importance of NAEP hope NAGB will carefully assess the benefits and costs of its new involvement in this area and inform Congress when it disagrees with proposed plans to follow this new course of action (especially if adequate financial and staff resources are not forthcoming).

8. What Role Should NAEP and NAGB Play in Improving American Education?

As we have seen, NAEP has been a useful tool for educational reform—letting us know how well students are doing in our schools and whether we have made progress over time. From fiscal year 1969 to fiscal year 1997 we spent $447 million (in constant 1996 dollars) on NAEP—a substantial but important investment that should be continued and even increased in the future.[326] Yet from the broader perspective of educational reform, which the Alexander-James report and its panel of NAE commentators recommended, we have not always developed or used NAEP optimally; nor will NAEP by itself be sufficient to achieve the educational reforms and goals that we need.

Much of the rationale for federal investment in major categorical educational programs, such as Title I of the Elementary and Secondary Education Act of 1965, has been to help disadvantaged students, especially those from families living in poverty. Yet for the past 30 years NAEP has not provided adequate information on the parental income or wealth of students. Asking about whether a student participates in the federally subsidized school lunch program is a helpful, but still inadequate index of a family's economic well-being—especially among students in junior and senior high school.[327] NCES, the agency with primary responsibility for developing NAEP background information, has taken some useful steps in the right direction, but now must work even harder to resolve this issue. NAGB, which has sometimes expressed concerns about the inclusion of certain types of background questions, needs to help NCES finally move forward on this matter.[328] Otherwise we are seriously limiting the analytic and reporting value of NAEP results and presenting the American public and policymakers with misleading information about the impact of such variables as race and ethnicity on student achievement.[329]

Similarly, NCES needs to continue its efforts to upgrade the quality of the statistical analyses that it provides using NAEP data. For decades, most of the reporting of NAEP results has been based only on simple descriptive statistics or on a cross-tabulation of the data. During the past 30 years, statistical analysis in the social sciences has increasingly relied on the use of multivariate techniques that can and should be employed in analyzing student achievement data. Recent work at NCES reveals the ability and willingness of that agency to use more sophisticated techniques of analysis, but NCES might still need additional encouragement to move more quickly in this direction.[330]

NAGB and NCES should also work together to develop more innovative and sophisticated ways of using NAEP data in other studies and investigations. So much time and effort have been spent in the last decade on developing state-level NAEP or setting student performance standards that not enough resources and thought have gone into exploring other ways in which NAEP might be used to analyze and improve American educational development. Given the large amount of funds we spend on NAEP annually, as well as hoped-for future increases, we need to work harder to see what uses we can make of the NAEP results in addition to being a very important, but still limited index of student educational achievement and progress. If NAGB and NCES do not have the interest or staff to explore these matters now, perhaps some outside group such as NAE or NAS could be persuaded to help, or another Alexander-James-type commission could be created to study the broader relationship between NAEP and American educational reform.

Finally, we should recognize and acknowledge that neither NAEP nor even the national voluntary tests will tell us how to reorganize our schools or directly improve classroom practices. After spending more than $150 billion on federal compensatory education programs in the past 30 years, we still don't know which programs work most effectively in different settings.[331] Robert Slavin aptly summarized the situation:

> For decades, policymakers have complained that the federal research and development enterprise has had too little impact on the practice of education. With few notable exceptions, this perception is, I believe, largely correct. Federally funded educational R&D has done a good job of producing information to inform educational practice, but has created few well-validated programs or practices that have entered widespread use. The limited direct influence of federal R&D compared to that of, say, research in medicine, physics, and chemistry can certainly be ascribed in part to the far more limited federal investment in educational R&D coupled with federal policies opposing investment in curriculum development dating back to the Nixon administration and a conservative backlash against such values-laden curricula as *Man: A Course of Study* in the 1970s.[332]

Congress and the administration need to pay the same kind of attention to reforming and improving educational research and development that they have done in overseeing the work of NAGB and NCES in regard to NAEP. For example, after spending approximately $1.5 billion on the regional educational laboratories and $1.1 billion on the research and development centers (in constant 1996 dollars) since the mid-1960s, why don't we have the necessary research and development knowledge to make the school improvements we need?[333] And why hasn't the Policy and Evaluation Service in the U.S. Department of Education provided the large-scale, rigorous, comparative evaluations of alternative educational programs necessary to improve the delivery of education to our children in different settings and circumstances? [334]

As I recently stated elsewhere:

> When existing federal educational programs, well-intentioned though they may be, are not as effective as they could or should be, the problem is not just wasted tax dollars, but wasted chances to help those in need. We raise the expectations of those who have the least to look forward to and then dash their hopes by failing to really help them escape from their poverty. The overall experiences with Title I and Head Start also have been frustrating for the American public, who have been willing to sacrifice for the achievement of the lofty goals of Title I and Head Start, but now find that little real progress has been made. For many of the at-risk students who pass through these programs and who are not significantly helped, however, the results are more than just frustrating—they are precious opportunities lost forever.[335]

NAGB and NAEP have certainly played an important role in telling us how well our children are doing in school and defining what our expectations of them should be. The next step, however, is to provide students with the type of effective education they need to reach the goals that we have set for them—something we still have not managed to accomplish. Perfecting the operation of NAGB and NAEP without simultaneously addressing the need for better research and development in the area of school improvement models and classroom practices ultimately makes little sense. Thus we need to go back to the broader visions of educational change envisioned in the original Alexander-James report to make NAEP an even more important and effective component of improving schools for all children in America.

Endnotes

[1] There are no in-depth, scholarly studies of the history of the National Assessment Governing Board. For a useful, 10-page history of NAGB, see Roy E. Truby, "National Assessment Governing Board: History and Accomplishments, 1988–1994," unpub. NAGB paper, March 21, 1994.

[2] James Axtell, *The School Upon a Hill: Education and Society in Colonial New England* (New Haven, CT: Yale University Press, 1974); Lawrence A. Cremin, *American Education: The Colonial Experience, 1607–1783* (New York: Harper and Row, 1970); Gerald F. Moran and Maris A. Vinovskis, *Religion, Family, and the Life Course: Explorations in the Social History of Early America* (Ann Arbor, MI: University of Michigan Press, 1992).

[3] Carl F. Kaestle, *Pillars of the Republic: Common Schools and American Society* (New York: Hill and Wang, 1983); Linda K. Kerber, *Women of the Republic: Intellect and Ideology in Revolutionary America* (Chapel Hill, NC: University of North Carolina Press, 1980); Mary Beth Norton, *Liberty's Daughters: The Revolutionary Experience of American Women, 1750–1800* (Boston: Little Brown, 1980).

[4] Samuel Bowles and Herbert Gintis, *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life* (New York: Basic Books, 1976); Lawrence A. Cremin, *American Education, the National Experience, 1783–1876* (New York: Harper and Row, 1980); Carl F. Kaestle and Maris A. Vinovskis, *Education and Social Change in Nineteenth-Century Massachusetts* (Cambridge: Cambridge University Press, 1980); Lee Soltow and Edward Stevens, *Rise of Literacy and the Common School in the United States: A Socio-Economic Analysis to 1870* (Chicago: University of Chicago Press, 1981).

[5] Kaestle, *Pillars of the Republic*; Maris A. Vinovskis, *Education, Society, and Economic Opportunity: A Historical Perspective on Persistent Issues* (New Haven, CT: Yale University Press, 1995).

[6] For a discussion of how Horace Mann used statistical information to further his educational efforts, see Charles L. Glenn, *The Myth of the Common School* (Amherst, MA: University of Massachusetts Press, 1988); Vinovskis, *Education, Society, and Economic Opportunity*.

[7] On the efforts to collect state and local educational data by national education associations in the 1830s and 1840s, see Donald R. Warren, *To Enforce Education: A History of the Founding Years of the United States Office of Education* (Detroit: Wayne State University Press, 1974), pp. 25–57.

[8] The best book on the origins of the U.S. Bureau of Education is Warren, *To Enforce Education*. See also the synthesis by Stephen Sniegoski, "History of the Department of Education and Its Forerunners," unpub. manuscript (July 1995). For an uncritical and laudatory treatment of Barnard's role, see Harry Kursh, *United State Office of Education: A Century of Service* (Philadelphia: Chilton Books, 1965), pp. 3–18.

[9] Sol Cohen, *Education in the United States: A Documentary History*, vol. 3, (New York: Random House, 1974), p. 1406.

[10] Warren, *To Enforce Education*; MacMullen, Barnard's most recent, sympathetic biographer, nevertheless concluded that: "In summary then, the failure of his first year as commissioner, for it is not unjust to label it a failure, was not as much the muddle of the

office, nor the sparseness of the data collected, nor the inability to realize the expectations of his supporters, nor even the political turmoil he generated. Rather it was Henry Barnard's own failure, and it was seated deep in his very being." Edith MacMullen, *In the Cause of True Education: Henry Barnard and Nineteenth-Century School Reform* (New Haven, CT: Yale University Press, 1991), p. 270.

[11] For a helpful, but mainly descriptive account of the agency in the early twentieth century, see Darrell H. Smith, *The Bureau of Education: Its History, Activities, and Organization* (Baltimore: Johns Hopkins University Press, 1923). On the educational efforts in Alaska, see Victor W. Hennigsen, III, "Reading, Writing and Reindeer: The Development of Federal Education in Alaska, 1877–1920," unpub. Ed.D. diss. (Harvard University, 1987).

[12] David S. Webster, "The Bureau of Education's Suppressed Rating of Colleges, 1911–1912," *History of Education Quarterly*, 24(3)(Winter 1984):499–510.

[13] Diane Ravitch, *The Troubled Crusade: American Education, 1945–1980* (New York: Basic Books, 1983); Maris A. Vinovskis, "The Changing Role of the Federal Government in Educational Research and Statistics," *History of Education Quarterly*, 36(2) (Summer 1996):111–128.

[14] Robert A. Divine, *The Sputnik Challenge: Eisenhower's Response to the Soviet Satellite* (New York: Oxford University Press, 1993); Peter B. Dow, *Schoolhouse Politics: Lessons from the Sputnik Era* (Cambridge, MA: Harvard University Press, 1991); Richard A. Dershimer, *The Federal Government and Education R&D* (Lexington, MA: Lexington Books, 1976). The cooperative research program had been created in 1954 as P.L. 83–531.

[15] On the passage of ESEA and Johnson's educational programs, see Stephen K. Bailey and Edith K. Mosher, *ESEA: The Office of Education Administers a Law* (Syracuse, NY: Syracuse University Press, 1968); Hugh Davis Graham, *The Uncertain Triumph: Federal Education Policy in the Kennedy and Johnson Years* (Chapel Hill, NC: University of North Carolina Press, 1984); Philip Meranto, *The Politics of Federal Aid to Education in 1965: A Study in Political Innovation* (Syracuse, NY: Syracuse University Press, 1967).

[16] On the changes in the nature of evaluation in the federal government in the 1960s, see Laurence E. Lynn, Jr., *Public Management as Art, Science, and Profession* (Chatham, NJ: Chatham House, 1996); Arnold J. Meltsner, *Policy Analysts in the Bureaucracy* (Berkeley, CA: University of California Press, 1976).

[17] Francis Keppel, *The Necessary Revolution in American Education* (New York: Harper and Row, 1966), pp. 108–109.

[18] The best discussion of the role and intentions of Keppel in creating a national student assessment is found in James A. Hazlett, "A History of the National Assessment of Educational Progress, 1963–1973," unpub. Ed.D. diss. (University of Kansas, 1973). Several other works also analyze Keppel's role in the national assessment, including Linda Hummel Fitzharris, "An Historical Review of the National Assessment of Educational Progress from 1963 to 1991," unpub. Ed.D. diss. (University of South Carolina, 1993) and Ellen Condliffe Lagemann, *The Politics of Knowledge: The Carnegie Corporation, Philanthropy, and Public Policy* (Middletown, CT: Wesleyan University Press, 1989), pp. 213–215. On the close working relationship between Keppel and President Johnson on educational matters, see Elizabeth Christine Dietrich, "Francis Keppel and Lyndon Johnson, Unique Collaborators in the Struggle for Federal Aid: ESEA, 1965," unpub. Ed.D. diss. (Lehigh University, 1994).

[19] There is widespread agreement about the crucial role that Tyler played in the creation of NAEP. For example, one overview of the development of NAEP assessed Tyler's contribution to the creation of NAEP as follows: "The remaining influence—perhaps 'phenomenon' would be a better word—that shaped educational assessment in its first years was the

unique leadership of Ralph Tyler in formulating the concept of assessment and in winning support for, and organizing, a national assessment. With the advantage of both a wide involvement in the field of education and a deep understanding of the intellectual processes by which curricular objectives and criteria of evaluation are arrived at, he was able to obtain broad agreement on the form and content of a national assessment. Moreover, he had the detailed knowledge of item construction and test administration needed to organize effectively the difficult work of creating the assessment instrument and putting it into the field." R. Darrell Bock, Robert Mislevy, and Charles Woodson, "The Next Stage in Educational Assessment," *Educational Researcher* 11(3) (March 1982):5.

[20] ECAPE was renamed and slightly reorganized in 1968 as the Committee on Assessing the Progress of Education (CAPE). This signaled the change from the original planning group to the one that was supposed to oversee the assessment program.

[21] American Association of School Administrators and the National Education Association, *National Assessment: Pro and Con* (Washington, DC: author, 1966), pp. 16–17.

[22] For example, at the December 1963 and January 1964 Carnegie-funded conferences, it was decided after considerable discussion that the results would be released at the state as well as the national level. Hazlett, "A History of the National Assessment of Educational Progress, 1963–1973," pp. 34–51.

[23] Quoted in Richard A. Dershimer, *The Federal Government and Educational R&D* (Lexington, MA: Lexington Books, 1976), p. 117.

[24] American Association of School Administrators and the National Education Association, *National Assessment: Pro and Con*, p. 25.

[25] The Education Commission of the States (ECS) had been created in June 1966. It was a compact of states and had also been funded by the Carnegie Corporation.

[26] On the extended battles over the national assessment of students in the second half of the 1960s, see Fitzharris, "An Historical Review of the National Assessment of Educational Progress from 1963 to 1991;" Hazlett, "A History of the National Assessment of Educational Progress, 1963–1973."

[27] The Carnegie Corporation spent $2,432,900 on the project. Lagemann, *The Politics of Knowledge*, p. 214.

[28] For a detailed discussion of the development process, see Hazlett, "A History of the National Assessment of Educational Progress, 1963–1973."

[29] The ten curriculum areas initially designated for assessment were art, career and occupational development, citizenship, literature, math, music, reading, science, social studies, and writing. The initial plans called for assessing two subjects per year, repeated in five-year cycles. Funding problems in the early 1970s forced changes in the schedule. Fitzharris, "An Historical Review of the National Assessment of Educational Progress from 1963 to 1991," pp. 61–63.

Rather than focusing on grades, which might have more of an educational policy orientation but contained children of different ages, NAEP was designed to look at four age groups: 9-year-olds, 13-year-olds, 17-year-olds (in and out of school), and young adults ages 26 to 35. These age groups were selected in part because they roughly corresponded to the completion of primary, intermediate, secondary, and postsecondary education. National Assessment of Educational Progress, *Mathematics Technical Report: Exercise Volume* (Washington, DC: U.S. Government Printing Office, 1977), p. 1.

Unlike most tests, which try to find items that best help to differentiate among students, the philosophy behind NAEP was to develop questions that reflected different levels of competence among all students. *Educational Evaluation: Classic Works of Ralph W. Tyler*, eds., George F. Madaus and Daniel Stufflebeam (Boston: Kluwer Academic Publishers, 1989), pp. 227–237.

[30] Bock, Mislevy, and Woodson, "The Next Stage in Educational Assessment."

[31] The four geographic regions were Northwest, Central, West, and Southwest; the seven types of community were extreme rural, low metropolitan, high metropolitan, big city, fringes around big cities, medium city, and smaller places. National Assessment of Educational Progress, *Citizenship: An Overview, 1975–76* (Denver: Education Commission of the States, 1978), p. 57.

[32] Parental education was defined by the highest level attained by either parent and subdivided into three groups: not graduated from high school, graduated high school, and some post-high school education. National Assessment of Educational Progress, *Citizenship: An Overview, 1975–76*, p. 57.

[33] Initially during the planning phase, there had been discussions about the need to include data on the income of parents, but that proposal was dropped in the final phases because of fears that it might discourage parents and teachers from participating in the assessment as well as the difficulty of obtaining such information reliably from younger children. At one point, the planning group had proposed two socioeconomic background levels—rich and poor—divided along some dimension of poverty. Martin T. Katzman and Ronald S. Rosen, "The Science and Politics of National Educational Assessment," *The Record*, 71(4) (May 1970):571–586.

[34] Katzman and Rosen, "The Science and Politics of National Educational Assessment," p. 584.

[35] Galen Saylor, "National Assessment: Pro and Con," *The Record*, 71(4) (May 1970):596.

[36] Hazlett, "A History of the National Assessment of Educational Progress, 1963–1973."

[37] Fitzharris, "An Historical Review of the National Assessment of Educational Progress from 1963 to 1991," pp. 49–50.

[38] As quoted in Hazlett, "A History of the National Assessment of Educational Progress, 1963–1973," p. 294.

[39] As quoted in Dershimer, *The Federal Government and Educational R&D*, p. 117.

[40] U.S. General Accounting Office, *The National Assessment of Educational Progress: Its Results Need to be Made More Useful* (HRD–76–113; Washington, DC: U.S. Government Printing Office, 1976).

[41] Kenneth Earl Benton, "The Impact of National Assessment of Educational Progress on Educational Decision Making at the Legislative Level as Perceived by Commissioners of the Education Commission of the States," unpub. Ed.D. diss. (Duke University, 1979).

[42] For a useful discussion and analysis of the value of NAEP to policymakers, see Penny A. Sebring and Robert F. Boruch, "How is National Assessment of Educational Progress Used? Results of an Exploratory Study," *Educational Measurement: Issues and Practice* 2(1) (Spring 1983):16–20.

[43] U.S. Public Law 95–561 (November 1, 1978).

[44] U.S. Public Law 98–511 (October 19, 1984).

[45] Terrell Bell, *The Thirteenth Man: A Reagan Cabinet Memoir* (New York: Free Press, 1988); Ellen Nobuko Horiuchi, "The United States Department of Education, 1981–1985: Agenda for Abolishment," unpub. Ph.D. diss. (Arizona State University, 1990).

[46] For example, the Heritage Foundation did not attack the efforts of the federal government to collect and disseminate educational data and information. Onnalee McGraw, "Department of Education," in *The First Year*, ed. Richard N. Holwill (Washington, DC: Heritage Foundation, 1982), pp. 73–82.

[47] The data on NAEP allocations have been assembled from a variety of sources, including records from the hearings of the Senate and House appropriations committees. If the expenditures on NAEP are transformed into constant 1996 dollars, the long-term decrease in funding becomes even more apparent—from a high of $22.5 million in FY 1972 to $6.7 million in FY 1981.

[48] U.S. General Accounting Office, *Education Information: Changes in Funds and Priorities Have Affected Production and Quality*, GAO/PEMD–88–4 (Washington, DC: U.S. Government Printing Office, 1987); Deborah A. Verstegen, "Educational Fiscal Policy in the Reagan Administration," *Educational Evaluation and Policy Analysis*, 12(4) (Winter 1990):355–373; Deborah A. Verstegen and David L. Clark, "The Diminution in Federal Expenditures for Education During the Reagan Administration," *Phi Delta Kappan*, 70(2) (October 1988): 134–38.

[49] For a useful discussion of the historical background of the educational standards leading to the developments of the early 1980s, see Diane Ravitch, *National Standards in American Education: A Citizen's Guide* (Washington, DC: Brookings Institution, 1995), pp. 33–58. Some analysts have argued that educational reformers have deliberately exaggerated the weakness of education in America. For example, see David C. Berliner and Bruce J. Biddle, *The Manufactured Crisis: Myths, Fraud, and the Attack on America's Public Schools* (Reading, MA: Addison-Wesley, 1995).

[50] National Commission on Excellence in Education, *A Nation at Risk: The Imperative of Educational Reform* (Washington, DC: U.S. Government Printing Office, 1983).

[51] Bell, *The Thirteenth Man*, pp. 114–43.

[52] Alan L. Ginsburg, Jay Noell, and Valena White Plisko, "Lessons from the Wall Chart," *Educational Evaluation and Policy Analysis*, 10(1) (Spring 1988):1.

[53] For example, see statement of William J. Bennett, "Fourth Annual Wall Chart of State Education Statistics," U.S. Department of Education, February 10, 1987.

[54] For example, Robert C. Seaver, vice president of the College Board, said, "We feel it is completely inappropriate to use the SAT as a measure of comparison between states, school systems, or schools." Thomas Toch, "E.D. Issues Study Ranking States on Education," *Education Week* (January 11, 1984), retrieved from http://www.edweek.org/htb.

[55] Ginsburg, Noell, and Plisko, "Lessons from the Wall Chart," pp. 1–12.

[56] Thomas Toch, "School Chiefs Endorse State-By-State Comparisons," *Education Week* (November 11, 1984), retrieved from http://www.edweek.org/htb.

[57] As the movement toward testing for minimal levels of student proficiency gained momentum, more NAEP-like state tests were often abandoned due to the lack of funds. Thomas Toch, "Tests Don't Help Teachers Teach, Officials Argue," *Education Week* (June 26, 1982), retrieved from http://www.edweek.org/htb.

[58] Thomas Toch, "N.A.E.P. Ready to Report State Scores," *Education Week* (April 11, 1984), retrieved from http://www.edweek.org/htb; "National News Roundup," *Education Week* (October 24, 1984), retrieved from http://www.edweek.org/htb.

[59] Thomas H. Fisher, administrator of the student assessment division in Florida, observed that "there is some feeling that the states have the constitutional authority for dealing with

education and not the federal government—and NAEP is viewed, in some circles, as an arm of the federal government." Cindy Currence, "Plan for State-by-State Comparison of Scores Meets Resistance," *Education Week* (August 22, 1984), retrieved from http://www.edweek.org/htb.

[60] Southern Regional Educational Board, *Proceedings: 1984 Annual Meeting, Southern Regional Education Board, June 20–22, 1984, Boca Raton, Florida* (Atlanta, GA: Southern Regional Education Board, 1984), p.1.

[61] Southern Regional Educational Board, *Proceedings: 1984 Annual Meeting*, p. 3.

[62] Southern Regional Education Board, *Measuring Student Achievement: Comparable Test Results for Participating SREB States, the Region, and the Nation* (Atlanta, GA: Southern Regional Education Board, 1986).

[63] J.R. Sirkin, "Nation's School Chiefs Complete Plans to Gauge States, Students," *Education Week* (November 27, 1985), retrieved from http://www.edweek.org/htb.

[64] Thomas Toch, "National Assessments Called Ineffectual in Study," *Education Week* (January 12, 1982), retrieved from http://www.edweek.org/htb.

[65] Sheppard Ranbom, "E.T.S. Revises National Assessments to Make Them 'More Useful,'" *Education Week* (October 19, 1983), retrieved from http://www.edweek.org/htb.

[66] Samuel Messick, "Response to Changing Assessment Needs: Redesign of the National Assessment of Educational Progress," *American Journal of Education*, 94(1) (November 1985):90–105.

[67] Doyle accepted that NAEP should be financed by the federal government to measure national education progress. He also believed that the federal government and the states should collaborate on research and development, using the data from NAEP, to provide a more analytic and in-depth diagnosis of education. But he felt that prescriptions of how schools were operated and reformed should be left to the states and local governments with minimal control or mandates from the federal government. Doyle proposed three goals for the governance of NAEP: "To insulate the NAEP from undue federal control, its constitution should embody three objectives: 1) the NAEP should be free from transient political influence (although it should be subject to long-term changes in public policy), 2) the governance of the NAEP should incorporate provisions to insure continuity and 'institutional memory,' and 3) the governance of the NAEP should reflect the needs of its various clients (not to be confused with interest groups or constituencies)." Denis P. Doyle, "Federal Funds and State Interests: A Proposal for Governing the NAEP," *Phi Delta Kappan*, 64(9) (May 1983):645.

[68] Doyle, "Federal Funds and State Interests," p. 645.

[69] "The governing board would not be likely to include statisticians, psychometricians, and experts in the area of survey research. To insure the highest levels of technical competence, the governing board of the NAEP would have to assemble a standing technical advisory board, made up of recognized authorities in the field." Doyle, "Federal Funds and State Interests,"645.

[70] *Creating a Center for Education Statistics: A Time for Action*, ed., Daniel B. Levine, (Washington, DC: National Academy Press, 1986), p. 4.

[71] For a discussion of the developments in NCES after 1988, see Maris A. Vinovskis, "Changing Federal Strategies for Supporting Educational Research, Development, and Statistics," unpub. background paper prepared for OERI's National Educational Research and Policy Priorities Board, January 1998.

[72] "The 'nation's report card,' as NAEP is now often called, is the most prestigious, most extensive, and most valuable of our assessment efforts, and the high-quality data it gathers

and publishes are widely used in research, planning, and school administration." Lamar Alexander and H. Thomas James, *The Nation's Report Card: Improving the Assessment of Student Achievement. Report of the Study Group; with a Review of the Report by a Committee of the National Academy of Education* (Cambridge, MA: National Academy of Education, 1987), p. 3.

[73] Alexander and James, *The Nation's Report Card*, pp. 4–5.

[74] "And in every instance, the objective for content areas and the reporting and interpretation of results should be clear and meaningful to educators, policymakers, researchers, and the general public." Alexander and James, *The Nation's Report Card*, p. 8.

[75] Alexander and James, *The Nation's Report Card*, p. 8.

[76] Alexander and James, *The Nation's Report Card*, p. 38.

[77] "NAEP has sought to redress the criticism that it collects insufficient background information by enlarging the number of variables measured. The Study Group is now concerned, however, that the collection of greater volumes of background and contextual variables might entail the risk of lowering the rate of participation among schools and students. Of even greater concern are the questions of how relevant the present background and contextual variables are to matters of educational research and policy, and how readily this information can be translated into meaningful action toward the improvement of education.

The collection of school variables promises to improve substantially the value of the assessment as a policy research tool, and the Study Group recommends that this information be gathered on a regular basis. The Study Group strongly recommends, however, that school variables collected by the national assessment be demonstrated by previous, smaller-scale studies to have significant effects on student achievement. Otherwise, there are real dangers of making an already large and complex project even more formidable to data analysts, of escalating collection and processing costs, and of reducing participation of schools owing to the added response burden. It is crucial then that the assessment should not try to play an exploratory role with respect to variables affecting achievement, and the new Educational Assessment Council should see that it does not." Alexander and James, *The Nation's Report Card*, pp. 29–30.

[78] "The chief responsibility of the new council would be to shape each assessment, selecting the content areas to be tested, defining conceptually the ground covered in each area, setting test specifications, and identifying feasible achievement goals for each of the age and grade levels to be tested. Closely linked with this activity would be the EAC's responsibility to set policy—on such matters as maintaining the continuity over time of the assessment's data banks, setting standards and procedures for the use of test data, and arranging for periodic review of the whole assessment effort by outside bodies (such as the present Study Group). In addition the council would explain, interpret, and represent the national assessment to educators, government officials, and the general public, assisting them in their use of its invaluable resources.

In carrying out these tasks, the council, supervising a permanent staff of perhaps ten professionals, would work with state and local curriculum specialists and test directors, as well as other experts in each field of study. Advisors would be constituted in specialized panels, addressing such issues as selection of educational objectives, construction of assessment exercises, sampling procedures, reporting and interpreting student performance, and the potentials and limitations of educational assessments. As the national assessment expands its responsibilities to state and local educational authorities, advisory panels should also provide guidelines for state and local linkages with the national data, including standards for fair and constructive comparisons, and performance expectations. Each panel would be charged with providing the council with the technical knowledge required for policy decisions." Alexander and James, *The Nation's Report Card*, pp. 32–33.

[79] Alexander and James, *The Nation's Report Card*, pp. 33–34.

[80] "The federal government will select, fund and monitor a contractor to conduct the national assessment. The contractor would be responsible for all item development, test and questionnaire administration, processing, and analysis, and would release findings to the states and the federal government, disseminate them, and provide technical assistance to states and local districts who wished to link their assessments with national data. The test contractor would also respond to state and local requests for additional assessment data collections to be paid for by states and localities. All of these responsibilities would be conducted in compliance with the general policies, content areas, and objectives prescribed by the Educational Assessment Council. The council would also specify, in connection with each content area or skill, whether it was to be tested on a state-by-state basis or nationally." Alexander and James, *The Nation's Report Card*, p. 33.

[81] Alexander and James, *The Nation's Report Card*, p. 36.

[82] "Several of the recommendations of the Alexander-James report are of obvious merit...and deserve strong endorsement. The expansion of NAEP to include regular assessments in science, history, civics, and geography in addition to the basic skills of reading, writing, and mathematics is clearly warranted. The change in design to facilitate the collection of data and reporting of results for individual states and localities is another important development. The commitment to continuity of future data collection with that of the past is essential to ensure the validity of comparisons over time." Alexander and James, *The Nation's Report Card*, p. 49.

[83] Alexander and James, *The Nation's Report Card*, p. 49.

[84] "For example, as we noted earlier, the basic NAEP design is not very useful for examining the effects of instructional variables on school outcomes. One option would be to support smaller, more intensive assessments specifically designed to address these institutional questions. This would permit a greater understanding of the variation among districts within states and among schools within districts. NAEP might select for further study specific schools that represent important pedagogical concerns—for example, institutions that are particularly strong in science and mathematics for girls'and minorities, or schools that accelerate learning for gifted children. In general, there are many opportunities where in-depth study of carefully selected sites would enable us to learn much about possible school improvements. Taking this approach, NAEP would provide not only an annual reporting on 'how we are doing' but also a constant flow of new knowledge on how various aspects of the schooling process and organization contribute to learning." Alexander and James, *The Nation's Report Card*, pp. 60–61.

[85] Alexander and James, *The Nation's Report Card*, p. 59.

[86] Alexander and James, *The Nation's Report Card*, pp. 59–60.

[87] Alexander and James, *The Nation's Report Card*, p. 60.

[88] "A great deal of test data is difficult to interpret. What does a level 400 on a reading test mean? Such scores can be used for comparison across time and localities, but the nation's report card could be more broadly informative if it provided clear descriptions of the levels of competence demonstrated by our children. Much more important than scale scores is the reporting of the proportions of individuals in various categories of mastery at specific ages." Alexander and James, *The Nation's Report Card*, p. 58.

[89] Alexander and James, *The Nation's Report Card*, p. 56.

[90] Quoted in Lynn Olson, "Bennett Panel Urges Major Expansion of NAEP," *Education Week* (March 25, 1987), retrieved from http://www.edweek.org/htb.

[91] "Chiefs Urge Changes in NAEP by 1990," *Education Week* (April 1, 1987), retrieved from http://www.edweek.org/htb.

[92] Quoted in Robert Rothman, "NAEP's Policy Board Endorses Redesign Plan, With Reservations," *Education Week* (June 17, 1987), retrieved from http://www.edweek.org/htb.

[93] Congressional Quarterly, *Almanac: 100th Congress, 1st Session....1987* (Washington, DC: Congressional Quarterly, 1988), pp. 521–31; Congressional Quarterly, *Almanac: 100th Congress, 2nd Session....1988* (Washington, DC: Congressional Quarterly, 1989), pp. 330–37.

[94] For example, during the lengthy testimony of Secretary Bennett before the House on the reauthorization of ESEA in March 1987, the issue of NAEP was not even raised. U.S. Congress, House Committee on Education and Labor, *Education and Training for American Competitiveness*, 100th Congress, 1st Session, March 11, 1987 (Washington, DC: U.S. Government Printing Office, 1987), pp. 417–83.

[95] U.S. House of Representatives, H.R. 5, *An Act to Improve Elementary and Secondary Education*, 100th Congress, 1st Session, June 3, 1987 (Washington, DC: U.S. Government Printing Office, 1987). Secretary Bennett opposed the successful efforts by Representative Visclosky in the House to add an amendment that would give NCES more independence: "More deeply objectionable are the establishment of a Presidentially-appointed Commissioner of Education Statistics and a separate authorization for the Center for Education Statistics (Center). In large measure, the Secretary reorganized the Office of Educational Research and Improvement (OERI) in 1985 to resolve conflicting lines of authority (posed principally by the Presidentially-appointed Director of the National Institute of Education and the policy-making National Council on Educational Research). In enacting Pub. L. 99–498 Congress, in effect, ratified the rationalized and streamlined management structure of the current OERI; the amendment, on the other hand, would reintroduce the type of bifurcated authority that led to the reorganization in the first place." Letter of Secretary William Bennett to Representative Augustus F. Hawkins, April 20, 1987, p. 13.

[96] The changes in the Center for Education Statistics were not intended to affect the earlier authorization provisions of NAEP, which had placed that program under the administrative jurisdiction of the Center. U.S. Congress, House Committee on Education and Labor, *Report on School Improvement Act of 1987*, 100th Congress, 1st Session, May 15, 1987 (Washington, DC: U.S. Government Printing Office, 1987), p. 98.

[97] "Section 9350 of the bill amends the authorization for the National Assessment of Educational Progress, under the General Provisions Act, to require this program (the primary source of national information about student achievement) to collect information on the performance of students served by Chapter 1. The National Assessment of Educational Progress has provided the Congress with important indicators of what is happening with the achievement of certain groups of students targeted for Chapter 1, but to date no information is specifically available on those children who are actually involved in Chapter 1." U.S. Congress, House Committee on Education and Labor, *Report on School Improvement Act of 1987*, p. 98.

[98] U.S. Congress, House Committee on Education and Labor, *Report on School Improvement Act of 1987*, p. 98.

[99] U.S. Senate, S. 373, *The Robert T. Stafford Elementary and Secondary Education Improvement Act of 1987*, 100th Congress, 1st Session, November 19, 1987 (Washington, DC: U.S. Government Printing Office, 1987).

[100] "The members of the Assessment Policy Committee, serving on the date of enactment of the National Assessment of Educational Progress Improvement Act, shall become members of the Board for the remainder of the terms of the appointment to the Assessment Policy Committee." The Secretary of Education was to appoint any other board members necessary to fill out the categories specified by the law. U.S. Congress, Senate, S. 373, *The Robert T. Stafford Elementary and Secondary Education Improvement Act of 1987*.

[101] "The Board shall be composed of—two Governors or former Governors, who shall not be members of the same political party; two State legislators, who shall not be members of the same political party; two chief State school officers; one superintendent of a local educational agency; one member of a State board of education; one member of a local board of education; two classroom teachers; one representative of business or industry; one curriculum planner or supervisor or testing and measurement expert; one nonpublic school administrator or policy maker; two school principals; one education researcher; and three additional members, who are representative of the general public, including parents. The Assistant Secretary for Educational Research and Improvement shall serve as an ex officio member of the Board as a nonvoting member." U.S. Congress, Senate, S. 373, *The Robert T. Stafford Elementary and Secondary Education Improvement Act of 1987*.

[102] U.S. Senate, S. 373, *The Robert T. Stafford Elementary and Secondary Education Improvement Act of 1987*.

[103] The Senate report language stated that "the staff provided to the NAGB are to represent the Board's interest to the Secretary of Education and to oversee the work of the contractor who administers the National Assessment." U.S. Senate, Committee on Labor and Human Resources, *The Robert T. Stafford Elementary and Secondary Education Improvement Act of 1987*, Report 100–222, 100th Congress, 1st Session, November 19, 1987 (Washington, DC: U.S. Government Printing Office, 1987), p. 89.

[104] "Groups Oppose Expansion of Federal Tests," *Education Week* (February 10, 1988), retrieved from http://www.edweek.org/htb.

[105] Julie A. Miller, "Conferees Likely to Loosen Rules in Bilingual Law," *Education Week* (March 16, 1988), retrieved from http://www.edweek.org/htb. The reference to the anomaly in the reading assessment was the fact that the large decline was probably due more to flaws in the testing instruments than to any such changes in the population being tested. Robert Rothman, "Drop in Scores on Reading Test Baffles Experts," *Education Week* (January 20, 1988), retrieved from http://www.edweek.org/htb; Robert Rothman, "NAEP's 'Anomaly' Blamed on Changes in Test's Design," *Education Week* (June 22, 1988), retrieved from http://www.edweek.org/htb.

[106] Julie A. Miller, "Accord Reached on Key Reauthorization Items," *Education Week* (April 16, 1988), retrieved from http://www.edweek.org/htb.

[107] Public Law 100–297.

[108] Public Law 100–297.

[109] Public Law 100–297.

[110] Public Law 100–297.

[111] The Senate bill had called for one curriculum planner or supervisor or testing and measurement expert as well as one education researcher. The final legislation called for two curriculum specialists as well as two testing and measurement experts. Public Law 100–297.

[112] Public Law 100–297.

[113] This prohibition against using NAEP scores to compare individuals, local schools, or local districts was in the legislation, but the idea that state-level results should not be used for ranking was not part of the law. As a result, the conference report on the states did not carry as much legal and political weight as if it had been written directly into the law. U.S. House of Representatives, *Elementary and Secondary Education: Conference Report to Accompany H.R. 5*, Report 100–567, 100th Congress, 2nd Session (Washington, DC: U.S. Government Printing Office, 1988), p. 387.

[114] Miller, "Accord Reached on Key Reauthorization Items," *Education Week* (April 16, 1988).

[115] Congressional Quarterly, *Almanac, 100th Congress, 2nd Session...1988* (Washington, DC: Congressional Quarterly, 1989), pp. 330–37.

[116] "Bennett Names Members of NAEP Board," *Education Week* (September 21, 1988), retrieved from http://www.edweek.org/htb. There was speculation that Bennett wanted to name the members of NAGB himself rather than leaving that responsibility to his successor, Lauro Cavazos. For discussions of Bennett's plans to leave the Reagan Administration, see Robert Rothman, "Discussing Rumors of a Rift, Bennett Sets September Exit," *Education Week* (May 18, 1988):1, 15; Julie A. Miller, "On Last Day, His Favorite Mix—Some Pedagogy, More Politicking," *Education Week* (September 28, 1988):1, 19.

[117] Public Law 100–297.

[118] Public Law 103–382.

[119] Letter from Secretary Richard W. Riley to William T. Randall, Chair, NAGB, April 14, 1995, available in NAGB briefing book (May 1995).

[120] Public Law 100–297.

[121] Public Law 103–382. Part of the impetus behind this change may have been to reduce the independence of the board somewhat by allowing the Secretary of Education to have more opportunity to change the composition of NAGB and to limit the number of experienced members on the board.

[122] When state Representative Wilhelmina F. Delco (D-TX) was not reappointed to the board, she charged that President Bush and his Secretary of Education "are interested in finding people agreeable to theirs." Robert Rothman, "New Members, Chairman Named to NAEP Board," *Education Week* (September 5, 1990), retrieved from http://www.edweek.org/htb.

[123] Rothman, "New Members, Chairman Named to NAEP Board."

[124] Calculated from a file created of the NAGB board members using a variety of records including summaries of the minutes available through NAGB.

[125] Lauro Cavazos was less interested in NAGB than either Alexander or Riley had been, but his Under Secretary Ted Sanders was a strong proponent of NAEP and NAGB.

[126] Public Law 100–297.

[127] There are many different ways in which regions are defined for analytic purposes. The present scheme is based on the divisions used in the American National Election studies. Warren E. Miller, *American National Election Study, 1988: Pre- and Post-Election Survey: Introduction and Codebook*, 2nd ed. (Ann Arbor, MI: Inter-University Consortium for Political and Social Research, 1990), pp. 8–10. This scheme follows most others except it includes Delaware in the Middle Atlantic States. The External States are Alaska and Hawaii, which in other configurations are sometimes left with the Pacific States.

[128] The census population figures for 1990 are calculated from U.S. Bureau of the Census, *Statistical Abstract of the United States: 1992* (112th ed.; Washington, DC: U.S. Government Printing Office, 1992), table numbers 25, 1339. Data on the regional residence of sixty-seven of the seventy board members were located—the other three cases were treated as missing data and not included in the tabulations.

[129] Daniel B. Taylor memo to Victor H. Ferry, Chair of the Nomination Committee, November 2, 1989, available in *NAGB Briefing Book* (December 1989).

[130] Daniel B. Taylor memo to members of the Nominating Committee, November 17, 1990, available in *NAGB Briefing Book* (December 1990).

[131] Board members received only token compensation for their efforts (besides reimbursement for travel expenses). Initially, board members received $100 for each day they worked specifically for NAGB.

[132] The data on attendance at meetings have been compiled from the minutes in the NAGB briefing books for November 1988 through March 1998. Data were not readily available for eight of the forty meetings—five of which occurred in 1989 and March 1990. Nevertheless, data were available for eighty percent of the meetings (or ninety-one percent of those since May 1990). While the results might have been slightly different if all of the data had been available, there is no reason to think that the overall results would be much different.

[133] There have been some differences in attendance rates for the different meetings. The special meetings attracted only a 72.5 percent attendance while the November meeting garnered an 87.7 percent attendance rate. The attendance rates for the other three regularly scheduled meetings were fairly similar—82.5 percent for March, 81.6 percent for May, and 78.5 percent for August.

[134] One might argue that the lower participation rate in the period 1995–98 is due in large part to the inclusion of two special board meetings, which had much lower rates of attendance. Therefore, the data for that period were recalculated without the presence of the two special meetings. This procedure, however, raised the participation rate only from 78.5 percent to 79.3 percent.

[135] Data are available on the classification of sixty-seven of the seventy members. A few members were moved from one category to another when they were reappointed. For purposes of this investigation, however, the category of the individual in the initial appointment was used throughout the analysis. The rates of participation were 92.7 percent for teachers, 92.2 percent for the general public, 93.1 percent for testing and measurement experts, and 93.5 percent for local school board members.

[136] Governors frequently designated someone else to attend on their behalf. These individuals could participate in the meetings, but were not allowed to vote.

[137] The six groups were curriculum specialists (88.4 percent), state superintendents (85.2 percent), state legislators (82.5 percent), state board of education members (81.3), nonpublic administrators or policymakers (84.4 percent), and business or industry representatives (88.9 percent).

[138] The three individuals singled out were Chester Finn, Mark Musick, and Herbert Walberg. "Democrats Seek to Prevent from Joining NAEP Panel," *Education Week* (October 12, 1988), retrieved from http//www.edweek.org/htb.

[139] All the quotes are from "Democrats Seek to Prevent from Joining NAEP Panel," *Education Week* (October 12, 1988).

BEST COPY AVAILABLE

[140] Secretary Cavazos replied to Hawkins and his colleagues that former Secretary Bennett had considered the possibility of a conflict of interest; Bennett had been assured by legal counsel that no conflict existed. NAGB minutes, November 18–19, 1988.

[141] NAGB minutes, November 18–19, 1988.

[142] NAGB minutes, Working Group I, December 16, 1988.

[143] Slightly less important were technical knowledge, oral presentation skills, writing skills, and executive skills. The least important was knowledge of federal government procedures and processes. "Assessment Criteria for Selection," *NAGB Briefing Book* (January 1989).

[144] U.S. Department of Education, "News" (April 26, 1989).

[145] Information on the current staff is available on the Web, http://www.nagb.org.

[146] Public Law 100-297.

[147] The financial data on NAEP and NAGB have been compiled from the NAGB briefing books and the congressional appropriations documents.

[148] U.S. House of Representatives, Subcommittee of the Committee on Appropriations, *Hearings on the Departments of Labor, Health and Human Services, Education, and Related Agencies, Appropriations for 1991*, 101st Congress, 2nd Session, Part 6, February 28, 1990 (Washington, DC: U.S. Government Printing Office, 1990), p. 1736.

[149] Public Law 103-382.

[150] Information about this rather unusual and threatening episode can be found in Tom Topuzes, chair, Ad Hoc Fact Finding Committee, "House Passed Funding Limitation for NAGB," *NAGB Briefing Book* (February 1991).

[151] In constant 1996 dollars, NAGB funding went from $3.34 million in FY 92 to $3.30 million in FY 98.

[152] Alexander and James, *The Nation's Report Card*, pp. 8–9.

[153] Alexander and James, *The Nation's Report Card*, p. 13.

[154] Alexander and James, *The Nation's Report Card*, p. 33.

[155] Alexander and James, *The Nation's Report Card*, pp. 34–35.

[156] Alexander and James, *The Nation's Report Card*, p. 56.

[157] Public Law 100-297.

[158] Public Law 100-297.

[159] Public Law 100-297. The pilot programs referred to the efforts to develop state-level assessments.

[160] Public Law 103-382.

[161] The six-month study cost NCES $250,000. Millicent Lawton, "Study Assails Waste in Oversight of NAEP," *Education Week* (June 19, 1996), retrieved from http//www.edweek.org/htb.

[162] KPMG Peat Marwick and Mathtech, "A Review of the National Assessment of Educational Progress: Management and Methodological Procedures," Final Report (June 4, 1996), p. I-1.

[163] Marwick and Mathtech, "A Review of the National Assessment of Educational Progress," p. V-1.

[164] Marwick and Mathtech, "A Review of the National Assessment of Educational Progress," pp. V-1 and V-2.

[165] Marwick and Mathtech, "A Review of the National Assessment of Educational Progress," p. V-4.

[166] Lawton, "Study Assails Waste in Oversight of NAEP."

[167] Pascal D. Forgione, Jr., "Peat Marwick Management Review of the National Assessment of Educational Progress (NAEP)," memo to Marshall Smith, Under Secretary, U.S. Department of Education (July 17, 1996), pp. 5–6.

[168] William T. Randall, "Comments on Draft Copy of NAEP Management Review by Peat Marwick," memo to Jeanne E. Griffith, Acting Commissioner, NCES (May 23, 1996).

[169] Lawton, "Study Assails Waste in Oversight of NAEP."

[170] Wayne C. Riddle, *National Assessment of Educational Progress: Background and Reauthorization and Issues*, Congressional Research Service Report No. 98-348 EPW (Washington, DC: Congressional Research Service, April 16, 1998), p. CRS-19.

[171] Riddle, *National Assessment of Educational Progress*, pp. CRS-18–CRS-20.

[172] For a discussion of the developments in the 1980s, see section C of this essay.

[173] NAGB, "Positions on the Future of the National Assessment," approved December 9, 1989, available in *NAGB Briefing Book* (March 1990).

[174] Robert Rothman, "In Protest, Pittsburgh Pulls Out of State-Level NAEP," *Education Week* (March 7, 1990), retrieved from http://www.edweek.org/htb.

[175] International Reading Association, "Reading Educators Opposed to Misguided Assessment," news release, May 18, 1990.

[176] Gary W. Phillips, "Benefits of State-by-State Comparisons," *Educational Researcher*, 20(3) (April 1991):18.

[177] Daniel M. Koretz, "State Comparisons Using NAEP: Large Costs, Disappointing Benefits," *Educational Researcher*, 20(3) (April 1991):20.

[178] Koretz, "State Comparisons Using NAEP," pp. 20–21.

[179] *New York Times* (October 1, 1989).

[180] Lynn Olson, "Governors Urge NAEP Expansion to Compare States," *Education Week* (February 13, 1991), retrieved from http://www.edweek.org/htb.

[181] For example, see National Education Goals Panel, *Building a Nation of Learners*, 1995 (Washington, DC: U.S. Government Printing Office, 1995).

[182] Ann Bradley, "Bingaman Bill Would Require Annual 'Report Cards' on Goals," *Education Week* (February 7, 1990), retrieved from http://www.edweek.org/htb.

[183] Public Law 102-62.

[184] National Council on Education Standards and Testing, *Raising Standards for American Education: A Report to Congress, the Secretary of Education, the National Education Goals Panel, and the American People* (Washington, DC: U.S. Government Printing Office, 1992), pp. 4–5.

[185] National Council on Education Standards and Testing, *Raising Standards for American Education*, p. 28.

[186] Robert Rothman, "Group Urges 'Hitting the Brakes' on National Test," *Education Week* (January 28, 1992), retrieved from http://www.edweek.org/htb.

[187] Linda Darling-Hammond, "National Standards and Assessment Policy Development: A Cautionary Tale," *American Journal of Education*, 102(4) (August 1994):480–81.

[188] George Bohrnstedt, Robert Glaser, and Robert Linn, *Assessing Student Achievement in the States: The First Report of the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment: 1990 Trial State Assessment* (Stanford, CA: National Academy of Education, 1992), p. xvii.

[189] Bohrnstedt, Glaser, and Linn, *Assessing Student Achievement in the States* (1992).

[190] George Bohrnstedt, Robert Glaser, and Robert Linn, *The Trial State Assessment: Prospects and Realities: The Third Report of the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment: 1992 Trial State Assessment* (Stanford, CA: National Academy of Education, 1993).

[191] Richard A. Boyd, letter to Robert Glaser, Chair, NAEP Evaluation Panel, January 13, 1992.

[192] "Capital Update," *Education Week* (May 5, 1993), retrieved from http://www.edweek.org/htb.

[193] "Capital Update," *Education Week* (May 19, 1993), retrieved from http://www.edweek.org/htb.

[194] Public Law 103–382.

[195] Debra Viadero, "E.D., NAEP Board Spar Over Plan to Limit Tests," *Education Week* (November 17, 1993), retrieved from http://www.edweek.org/htb.

[196] George Bohrnstedt, Robert Glaser, and Robert Linn, *Quality and Utility: The 1994 Trial State Assessment in Reading: The Fourth Report of the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment: 1994 Trial State Assessment in Reading* (Stanford, CA: National Academy of Education, 1996), p. xxix.

[197] Bohrnstedt, Glaser, and Linn, *Quality and Utility* (1996), pp. xxix–xxx.

[198] Work Group on Planning, "Mission Statement for Planning," available in *NAGB Briefing Book* (March 1995).

[199] National Assessment Governing Board, "Policy Statement on Redesigning the National Assessment of Educational Progress" (August 2, 1996), pp. 1–2. Available in NAGB briefing book (September 1996).

[200] National Assessment Governing Board, "Policy Statement on Redesigning the National Assessment of Educational Progress," p. 5.

[201] National Assessment Governing Board, "Policy Statement on Redesigning the National Assessment of Educational Progress," p. 7.

[202] National Assessment Governing Board, "Policy Statement on Redesigning the National Assessment of Educational Progress," p. 17.

[203] National Assessment Governing Board, "Policy Statement on Redesigning the National Assessment of Educational Progress," p. 4.

[204] Discussed at the National Center for Education Statistics, "Forum on the National Assessment of Educational Progress (NAEP) Redesign," OERI, Washington, D.C. (June 3, 1997).

[205] National Assessment Governing Board, "Setting Mathematics Achievement Levels for the National Assessment of Educational Progress," June 3, 1991.

[206] The board was frequently reminded by the staff of the difficulties involved in developing performance standards in a situation in which there was no agreement on procedures for setting achievement levels. For example, Roy Truby quoted to the board statements from publications by ETS as well as the National Center for Research on Evaluation Standards and Testing that indicated that there was little agreement on what achievement standards are and how they should be set. Transcript of NAGB meeting, March 4, 1994.

[207] Public Law 95-561.

[208] Alexander and James, *The Nation's Report Card*, p. 32.

[209] Alexander and James, *The Nation's Report Card*, p. 58.

[210] U.S. Senate, S. 373, *The Robert T. Stafford Elementary and Secondary Education Improvement Act of 1987*.

[211] U.S. House of Representatives, H.R. 5, *An Act to Improve Elementary and Secondary Education*.

[212] Public Law 100-297.

[213] Terry Hartle was unusually knowledgeable about NAEP because he had worked for ETS for ten years, the last two of which dealing with issues relating to NAEP. Transcript of NAGB meeting, May 13, 1989.

[214] Transcript of NAGB meeting, May 13, 1989.

[215] U.S. House of Representatives, *Report on School Improvement Act of 1978; U.S. Congress, House, Elementary and Secondary Education: Conference Report to Accompany H.R. 5, 1988*, Report 100-567; U.S. Senate, *The Robert T. Stafford Elementary and Secondary Education Improvement Act of 1987*, Report 100-222, 100th Congress, 1st Session (Washington, DC: U.S. Government Printing Office, 1987).

[216] Harold Howe, II, letter to Emerson Elliott, Director of the Center for Education Statistics (May 25, 1988).

[217] For a discussion of state activity in testing students and setting achievement standards, see Robert Rothman, "States Turn to Student Performance as New Measure of School Quality," *Education Week* (November 8, 1989), retrieved from http://www.edweek.org/htb.

[218] Peter Schmidt, "Better Student-Performance Data Needed, Panel Told," *Education Week* (November 1, 1989), retrieved from http://www.edweek.org/htb. Ann Bradley, "Bingaman Bill Would Require Annual 'Report Cards' on Goals." For a useful discussion of the congressional politics of national standards, see John F. Jennings, *Why National Standards and Tests? Politics and the Quest for Better Schools* (Thousand Oaks, CA: Sage Publications, 1998).

[219] *New York Times* (October 1, 1989).

[220] On the creation and activities of the National Goals Panel, see Ravitch, *National Standards in American Education; Rothman, Measuring Up: Standards, Assessment, and School Reform* (San Francisco: Jossey-Bass, 1995).

[221] "Text of Statement on Education Goals Adopted by Governors," *Education Week* (March 7, 1990), retrieved from http://www.edweek.org/htb.

[222] Transcript of NAGB meeting, January 27, 1989.

[223] Transcript of NAGB meeting, January 27, 1989.

[224] Transcript of NAGB meeting, January 27, 1989.

[225] Joe Nathan, "Goals for American Students: A Central Issue for the National Assessment Governing Board," available in *NAGB Briefing Book* (December 1989).

[226] Roy Truby, "Staff Paper on Setting Goals for the National Assessment," December 8, 1989, available in *NAGB Briefing Book* (December 1989).

[227] National Assessment Governing Board, *Bulletin: National Assessment of Educational Progress* (February 1990), p. 2. Robert Rothman, "NAEP Plan to Set Performance Goals Questioned," *Education Week* (January 31, 1990), retrieved from http://www.edweek.org/htb.

[228] National Assessment Governing Board, *Bulletin* (February 1990), p. 2.

[229] National Assessment Governing Board, *Bulletin* (February 1990), p. 2.

[230] National Assessment Governing Board, *Bulletin* (February 1990), p. 3.

[231] National Assessment Governing Board, "Setting Achievement Goals for the National Assessment of Educational Progress," February 19, 1990, p. 3. Available in *NAGB Briefing Book* (March 1990).

[232] National Assessment Governing Board, "Setting Achievement Goals for the National Assessment of Educational Progress," February 19, 1990, p. 5.

[233] National Assessment Governing Board, "Setting Mathematics Achievement Levels for the National Assessment of Educational Progress," June 3, 1991, p. 6.

[234] Robert Rothman, "NAEP to Create Three Standards for Performance," *Education Week* (May 23, 1990), retrieved from http://www.edweek.org/htb.

[235] Rothman, "NAEP to Create Three Standards of Performance."

[236] This brief description of the process is from National Assessment Governing Board, "Setting Mathematics Achievement Levels for the National Assessment of Educational Progress," June 3, 1991, pp. 7–12. For additional discussion of this process, see Robert Rothman, "NAEP May Delay Report Detailing State-Level Data," *Education Week* (November 28, 1990), retrieved from http://www.edweek.org/htb.

[237] Robert Rothman, "Educators Worry NAEP Levels May Hurt Instructions," *Education Week* (December 5, 1990), retrieved from http://www.edweek.org/htb.

[238] Rothman, "Educators Worry NAEP Levels May Hurt Instructions."

[239] Robert Rothman, "NAEP Board Urged to Delay Standards-Setting Plan," *Education Week* (January 16, 1991), retrieved from http://www.edweek.org/htb.

[240] Rothman, "NAEP Board Urged to Delay Standards-Setting Plan."

[241] For example, see Robert L. Linn, Daniel M. Koretz, Eva L. Baker, and Leigh Burstein, *The Validity and Credibility of the Achievement Levels for the 1990 National Assessment of Educational Progress in Mathematics* (Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing, January 1991).

[242] For a discussion of the work of these researchers, see Hummel, "An Historical Review of the National Assessment of Educational Progress from 1963 to 1991," pp. 178–92.

[243] Hummel, "An Historical Review of the National Assessment of Educational Progress from 1963 to 1991," pp. 178–92.

[244] Daniel L. Stufflebeam, Richard M. Jaeger, and Michael Scriven, "Summative Evaluation of the National Assessment Governing Board's Inaugural Effort to Set Achievement Levels of the National Assessment of Educational Progress," draft report submitted to NAGB, August 1, 1991, p. ii.

[245] Stufflebeam, Jaeger, and Scriven, "Summative Evaluation of the National Assessment Governing Board's Inaugural Effort," p. ii.

[246] Stufflebeam, Jaeger, and Scriven, "Summative Evaluation of the National Assessment Governing Board's Inaugural Effort," p. iii.

[247] Stufflebeam, Jaeger, and Scriven, "Summative Evaluation of the National Assessment Governing Board's Inaugural Effort," pp. 9–10.

[248] Richard Boyd, "Draft Summative Evaluation Report," memo to NAGB members, August 14, 1991.

[249] Richard Boyd, Mark Musick, and Michael Glode, letter to Edward Haertel, August 14, 1991.

[250] Hummel, "An Historical Review of the National Assessment of Educational Progress from 1963 to 1991," p. 184.

[251] National Assessment Governing Board, "Response to the Draft Summative Evaluation Report on the National Assessment Governing Board's Inaugural Effort to Set Achievement Levels on the National Assessment of Educational Progress," August 14, 1991.

[252] Hummel, "An Historical Review of the National Assessment of Educational Progress from 1963 to 1991," p. 184.

[253] Robert Rothman, "NAEP Board Fires Researchers Critical of Standards Process," *Education Week* (September 4, 1991), retrieved from http://www.edweek.org/htb.

[254] Rothman, "NAEP Board Fires Researchers Critical of Standards Process."

[255] General Accounting Office, *Educational Achievement Standards: NAGB's Approach Yields Misleading Interpretations* (GAO/PEMD–93–12; Washington, DC: U.S. Government Printing Office, June 1993), p. 1.

[256] General Accounting Office, *Educational Achievement Standards*, p. 3.

[257] General Accounting Office, *Educational Achievement Standards*, pp. 4–5.

[258] General Accounting Office, *Educational Achievement Standards*, p. 5.

[259] General Accounting Office, *Educational Achievement Standards*, p. 70.

[260] General Accounting Office, *Educational Achievement Standards*, p. 77.

[261] General Accounting Office, *Educational Achievement Standards*, p. 77.

[262] General Accounting Office, *Educational Achievement Standards*, p. 78.

[263] Robert Rothman, "G.A.O. Assails Standards-Setting Process for NAEP," *Education Week* (March 18, 1992), retrieved from http://www.edweek.org/htb.

[264] Debra Viadero, "G.A.O. Blasts Method for Reporting NAEP Results," *Education Week* (July 14, 1993), retrieved from http://www.edweek.org/htb.

[265] Viadero, "G.A.O. Blasts Method for Reporting NAEP Results."

[266] Mark Musick, letter to the editor, *Education Week* (August 4, 1993), retrieved from http://www.edweek.org/htb. The editor attached the following note to the end of Musick's

letter: "The phrase 'fundamentally flawed' appears in the draft report, not the final report. We regret the error."

[267] Lorrie Shepard, Robert Glaser, Robert Linn, and George Bohrnstedt, *Setting Performance Standards for Student Achievement: A Report of the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment: An Evaluation of the 1992 Achievement Levels* (Stanford, CA: National Academy of Education, 1973).

[268] Shepard, Glaser, Linn, and Bohrnstedt, *Setting Performance Standards for Student Achievement*, p. xxi.

[269] Shepard, Glaser, Linn, and Bohrnstedt, *Setting Performance Standards for Student Achievement*, p. xxi.

[270] Shepard, Glaser, Linn, and Bohrnstedt, *Setting Performance Standards for Student Achievement*, p. xxii.

[271] Shepard, Glaser, Linn, and Bohrnstedt, *Setting Performance Standards for Student Achievement*, pp. xxiii–xxiv.

[272] Shepard, Glaser, Linn, and Bohrnstedt, *Setting Performance Standards for Student Achievement*, p. xxiv.

[273] Shepard, Glaser, Linn, and Bohrnstedt, *Setting Performance Standards for Student Achievement*, pp. xxiv–xxix.

[274] Shepard, Glaser, Linn, and Bohrnstedt, *Setting Performance Standards for Student Achievement*, pp. xxix–xxx.

[275] Shepard, Glaser, Linn, and Bohrnstedt, *Setting Performance Standards for Student Achievement*, p. xxxii.

[276] Michael Kane, "Comments on the NAE Evaluation of the NAGB Achievement Levels," unpub. paper, September 1993, available in *NAGB Briefing Book* (November 1993).

[277] Gregory J. Cizek, "Reactions to National Academy of Education Report, 'Setting Performance Standards for Student Achievement,'" unpub. paper, August 1993, pp. 33–34; available in *NAGB Briefing Book* (November 1993).

[278] For example, the NAE report and the criticisms of it were extensively discussed at the November 1993 NAGB meeting.

[279] For example, the article in *Education Week* covering this issue gave much more coverage and credibility to the NAE panel's report than to the rebuttals by the evaluation critics, who were not even named in that essay. Debra Viadero, "Yet Another Report Assails NAEP Assessment Methods," *Education Week* (September 22, 1993), retrieved from http://www.edweek.org/htb.

[280] Linda Crocker and Michael Zieky, *Joint Conference on Standard Setting for Large-Scale Assessments, Executive Summary* (Washington, DC: U.S. Government Printing Office, 1994), Vol. I, p. ES-10.

[281] Larry Feinberg, "Concluding Summary of Joint Conference on Standard Setting for Large-Scale Assessments," distributed to NAGB, October 1994. Feinberg prepared a written draft of Popham's concluding statements, which then were checked and approved by Popham before distribution.

[282] Quoted in Feinberg, "Concluding Summary of Joint Conference on Standard Setting for Large-Scale Assessments."

[283] United States General Accounting Office, *Educational Achievement Standards*, pp. 66–67.

[284] The separation occurred in part because NCES had already prepared the math report before the achievement levels for the 1990 math assessments had been completed.

[285] The board held frequent discussions about the relationship between NCES and NAGB in the reporting of the achievement standards. For example, see the transcripts of the NAGB meetings of November 20, 1993, and March 5, 1994.

[286] Ina V.S. Mullis, Jay R. Campbell, and Alan E. Farstrup, *NAEP 1992 Reading Report Card for the Nation and the States* (NCES Report No. 23–ST06; Washington, DC: U.S. Government Printing Office, September 1993), pp. 49–50.

[287] NAGB continued to insist that the achievement standards be given a prominent place in the reporting of the NAEP results. On March 5, 1994, the board unanimously passed a resolution that stated: "Noting its resolution of November 20, 1993, which reaffirmed its policy that achievement levels shall be the initial and primary means of reporting NAEP results; Noting that this resolution provides a section on achievement levels shall be a prominent part of initial reports on 1994 NAEP reading results.... Therefore resolves that efforts be redoubled to reach agreement with NCES assuring that Board policy is carried out in the state-level NAEP reports." *NAGB Briefing Book* (May 1994).

[288] Jay R. Campbell, Patricia L. Donahue, Clyde M. Reese, and Gary W. Phillips, *NAEP 1994 Reading Report Card for the Nation and the States: Findings from the National Assessment of Educational Progress and Trial State Assessment* (Washington, DC: U.S. Government Printing Office, January 1996), p. 6.

[289] "Capital Digest," *Education Week* (May 19, 1993), retrieved from http://www.edweek.org/htb.

[290] Debra Viadero, "House Subcommittee Approves Proposal to Abolish NAEP," *Education Week* (February 9, 1994), retrieved from http://www.edweek.org/htb.

[291] Viadero, "House Subcommittee Approves Proposal to Abolish NAEP."

[292] Mark D. Musick, "Why NAEP Needs Its Independent Governing Board," *Education Week* (February 9, 1994), retrieved from http://www.edweek.org/htb.

[293] Quoted in Musick, "Why NAEP Needs Its Independent Governing Board."

[294] Education Information Advisory Committee of the Council of Chief State School Officers, "National Assessment of Educational Progress (NAEP) Reading Achievement Levels," Resolution passed at Alexandria, Virginia, May 4, 1994.

[295] Education Information Advisory Committee of the Council of Chief State School Officers, "National Assessment of Educational Progress (NAEP) Reading Achievement Levels."

[296] Mark Pitsch, "House Poised to Clear E.S.E.A. Reauthorization," *Education Week* (March 9, 1994), retrieved from http://www.edweek.org/htb.

[297] Public Law 103–382.

[298] Wayne C. Riddle, *National Assessment of Educational Progress: Background and Reauthorization Issues*, CRS Report 98–348 EPW (Washington, DC: Congressional Research Service, April 16, 1998), pp. CRS–6–CRS–7.

[299] Robert Glaser, Robert Linn, and George Bohrnstedt, *Assessment in Transition: Monitoring the Nation's Educational Progress* (Stanford, CA: National Academy of Education, 1997), p. xv.

[300] Millicent Lawton, "NAEP Assigns No Grades on Science Exam," *Education Week* (May 7, 1997), retrieved from http://www.edweek.org/htb.

[301] Christine Y. O'Sullivan, Clyde M. Reese, and John Mazzeo, *NAEP 1996 Science Report Card for the Nation and the States* (Washington, DC: National Center for Education Statistics, 1997).

[302] Mary Lyn Bourque, Audrey B. Champagne, and Sally Crissman, *1996 Science Performance Standards: Achievement Results for the Nation and the States* (Washington, DC: National Assessment Governing Board, 1997); Millicent Lawton, "Science Proves a Big Mystery To U.S. Pupils," *Education Week* (October 29, 1997), retrieved from http://www.edweek.org/htb.

[303] Public Law 103–382.

[304] Quoted in Millicent Lawton, "Board Approves Plans for NAEP Civics Tests," *Education Week* (March 13, 1996), retrieved from http://www.edweek.org/htb.

[305] On the role of history in educational policymaking, see Maris A. Vinovskis, *History and Educational Policymaking* (New Haven, CT: Yale University Press, 1999), pp. 239–56.

[306] This is true even though we have not progressed very far in agreeing on national content standards and therefore cannot align our assessments to them at this time.

[307] For example, see Glaser, Linn, and Bohrnstedt, *Assessment in Transition*.

[308] Alexander and James, *The Nation's Report Card*.

[309] National Center for Education Statistics, *The NAEP Guide: A Description of the Content and Methods of the 1997 and 1998 Assessments*, NCES 97–990, rev. ed. (Washington, DC: U.S. Government Printing Office, 1997), p. 73.

[310] Gary W. Phillips, "Technical, Methodological, and Operational Issues in the Implementation of the NAGB Redesign of NAEP," presentation at OERI, June 3, 1997.

[311] On the role of immigration in changing American society, see Roberto Suro, *Strangers Among Us: How Latino Immigration is Transforming America* (New York: Alfred Knopf, 1998); U.S. Commission on Immigration Reform, *Becoming an American: Immigration and Immigrant Policy, 1997 Report to Congress* (Washington, DC: U.S. Government Printing Office, 1997).

[312] Richard M. Battistoni, *Public Schooling and the Education of Democratic Citizens* (Jackson, MS: University Press of Mississippi, 1985); Joan N. Burstyn, ed., *Educating Tomorrow's Valuable Citizen* (Albany, NY: State University of New York Press, 1996); John Marciano, *Civic Illiteracy and Education: The Battle for the Hearts and Minds of American Youth* (New York: Peter Lang, 1997); Gary B.Nash, Charlotte Crabtree, and Ross E. Dunn, *History on Trial: Culture Wars and the Teachers of the Past* (New York: Alfred Knopf, 1997); Walter C. Parker, *Educating the Democratic Mind* (Albany, NY: State University of New York Press, 1996); Diane Ravitch and Chester E. Finn, Jr., *What Do Our 17-Year-Olds Know? A Report on the First National Assessment of History and Literature* (New York: Harper and Row, 1987); David T. Sehr, *Education for Public Policy* (Albany, NY: State University of New York Press, 1997).

[313] In general, the U.S. Department of Education has shown relatively little interest in history, geography, or civics and the other social sciences in its reform initiatives.

[314] Phillips, "Technical, Methodological, and Operational Issues."

[315] The federal government might fund entirely some of the state-level assessments, but the states should have more flexibility on how to spend their research and development

monies if those assessments are given on a more frequent basis. That would ensure that more frequent state assessments were the highest priority for the states, compared with alternate uses of that money.

[316] On the need for maintaining history as a discipline separate from social studies, see Paul Gagnon, ed., *Historical Literacy: The Case for History in American Education* (New York: MacMillan, 1989).

[317] On the changing role of the regional educational laboratories in educational development, see Maris A. Vinovskis, "Analysis of the Quality of Research and Development at the OERI Research and Development Centers and at the OERI Regional Educational Laboratories," final report submitted to OERI, June 1993.

[318] This purchase program might operate like a modified purchase program plan proposed for the labs in the late 1960s and early 1970s. Vinovskis, "Analysis of the Quality of Research and Development."

[319] On the ongoing debate over the need for national standards and assessments, see Diane Ravitch, ed., *Debating the Future of American Education: Do We Need National Standards and Assessments?* (Washington, DC: Brookings Institution, 1995).

[320] For discussions on what knowledge citizens might need to be involved effectively in American politics, see Michael X. Delli Carpini and Scott Ketter, *What Americans Know About Politics and Why It Matters* (New Haven, CT: Yale University Press, 1989); W. Russell Neuman, *The Paradox of Mass Politics: Knowledge and Opinion in the American Electorate* (Cambridge, MA: Harvard University Press,1986); Norman H. Nie, Jane Junn, and Kenneth Stehlik-Barry, *Education and Democratic Citizenship in America* (Chicago: University of Chicago Press, 1996).

[321] Only one Secretary of Education (Riley) has operated so far under the new selection procedures.

[322] Some NAGB staff members have published independently and continue to make useful contributions at professional meetings.

[323] On the roles of congressional professional staffs, see Harrison W. Fox, Jr., and Susan Webb Hammond, *Congressional Staff: The Invisible Force in American Lawmaking* (New York: Free Press, 1977).

[324] For a useful discussion and overview of the voluntary national tests, see Wayne Riddle, "National Tests: Administrative Initiative," *CRS Report for Congress*, 97-774, revised, EPW (June 16, 1998).

[325] For example, given the limited amount of time likely to be allocated to any single test and the lack of matrix sampling, if subjects like American history were tested, one could imagine situations in which students whose teachers did not emphasize a particular time period or subject would be strongly disadvantaged, even though they might know a considerable amount about other topics in history.

[326] The amount of money spent was calculated from data assembled for my project on educational research and development expenditures.

[327] Some needy students, especially those in high school and from certain cultural backgrounds, are less likely to participate in the school lunch program.

[328] Much of the debate about background questions is over whether NAEP scores should be adjusted for background characteristics before being released, not the issue of using background data for analytic purposes.

[329] For example, NAEP results for American history by race and ethnicity, without considering the socioeconomic background of the students, can easily lead to misinterpretations of the factors behind variations in student achievement levels.

[330] For recent NCES publications that employ multivariate techniques of analysis, see Paul T. Decker, Jennifer King Rice, and Mary T. Moore, *Education and the Economy: An Indicators Report*, NCES 97–269 (Washington, DC: U.S. Government Printing Office, 1997); Laura Lippman, Shelley Burns, and Edith McArthur, *Urban Schools: The Challenge of Location and Poverty*, NCES 96–864 (Washington, DC: U.S. Government Printing Office, 1996).

[331] Maris A. Vinovskis, "The Development and Effectiveness of Compensatory Education Programs," in *Giving Better, Giving Smarter: Working Papers of the National Commission on Philanthropy and Civic Renewal*, eds. John W. Barry and Bruno V. Mano (Washington, DC: National Commission on Philanthropy and Civic Renewal, 1997), pp. 169–92.

[332] Robert E. Slavin, "Design Competitions: A Proposal for a New Federal Role in Educational Research and Development," *Educational Researcher*, 26, No. 1 (January/February 1997):22.

[333] On the regional education laboratories and R&D centers, see Vinovskis, "Analysis of the Quality of Research and Development."

[334] Most large-scale evaluation at PES focuses on student outcomes without trying to ascertain whether a particular model or approach is working better than others.

[335] Prepared testimony of Maris Vinovskis for the U.S. House of Representatives, Subcommittee on Early Childhood, Youth and Families, October 2, 1998 (hearing has been postponed).

BEST COPY AVAILABLE